

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/10, 15/62, 15/63</b>		A1	(11) International Publication Number: <b>WO 97/07205</b> (43) International Publication Date: <b>27 February 1997 (27.02.97)</b>
<p>(21) International Application Number: <b>PCT/DK96/00343</b></p> <p>(22) International Filing Date: <b>12 August 1996 (12.08.96)</b></p> <p>(30) Priority Data: 0907/95 11 August 1995 (11.08.95) DK 1047/95 20 September 1995 (20.09.95) DK</p> <p>(71) Applicant (<i>for all designated States except US</i>): NOVO NORDISK A/S [DK/DK]; Novo Allé, DK-2880 Bagsværd (DK).</p> <p>(72) Inventor; and (75) Inventor/Applicant (<i>for US only</i>): OKKELS, Jens, Sigurd [DK/DK]; Novo Nordisk a/s, Novo Allé, DK-2880 Bagsværd (DK).</p> <p>(74) Agent: NOVO NORDISK A/S; Corporate Patents, Novo Allé, DK-2880 Bagsværd (DK).</p>		<p>(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b> <i>With international search report.</i></p>	
<p>(54) Title: METHOD FOR PREPARING POLYPEPTIDE VARIANTS</p> <p>(57) Abstract</p> <p>The present invention relates to a method for preparing positive polypeptide variants by shuffling different nucleotide sequences of homologous DNA sequences by <i>in vivo</i> recombination comprising the steps of a) forming at least one circular plasmid comprising a DNA sequence encoding a polypeptide, b) opening said circular plasmid(s) within the DNA sequence(s) encoding the polypeptide(s), c) preparing at least one DNA fragment comprising a DNA sequence homologous to at least a part of the polypeptide coding region on at least one of the circular plasmid(s), d) introducing at least one of said opened plasmid(s), together with at least one of said homologous DNA fragment(s) covering full-length DNA sequences encoding said polypeptide(s) or parts thereof, into a recombination host cell, e) cultivating said recombination host cell, and f) screening for positive polypeptide variants.</p>			

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

Despite the existence of the above methods there are still need for even better iterative *in vivo* recombination methods for preparing novel positive polypeptide variants.

5

#### SUMMARY OF THE INVENTION

The object of the present invention is to provide an improved method for preparing positive polypeptide variants by an *in vivo* 10 recombination method.

The inventor of the present invention have surprisingly found that such positive polypeptide variants may advantageously be prepared by shuffling different nucleotide sequences of homologous DNA sequences 15 by *in vivo* recombination comprising the steps of

- a) forming at least one circular plasmid comprising a DNA sequence encoding a polypeptide,
- b) opening said circular plasmid(s) within the DNA sequence(s) 20 encoding the polypeptide(s),
- c) preparing at least one DNA fragment comprising a DNA sequence homologous to at least a part of the polypeptide coding region on at least one of the circular plasmid(s),
- d) introducing at least one of said opened plasmid(s), together with at least one of said 25 homologous DNA fragment(s) covering full-length DNA sequences encoding said polypeptide(s) or parts thereof, into a recombination host cell,
- e) cultivating said recombination host cell, and
- f) screening for positive polypeptide variants.

30

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the yeast expression plasmid pJS026 comprising DNA sequence encoding the *Humicola lanuginosa* lipase gene.

35 Figure 2 shows the yeast expression plasmid pJS037, comprising DNA sequence encoding the *Humicola lanuginosa* lipase gene containing twelve additional restriction sites.

Figure 3 shows the plasmid pJS026.

Figure 4 shows the plasmid pJS037.

40 Figure 5 shows the *in vivo* recombination of the 0.9 kb synthetic wild-type *Humicola lanuginosa* lipase with pJS037 using *Saccharomyces*

*cerevisiae* as the recombination host cell (described in Example 1). Figure 6 shows the *in vivo* recombination of a DNA fragment prepared from *Humicola lanuginosa* lipase variant (y) with *Humicola lanuginosa* lipase variant (d) comprised in a plasmid using *Saccharomyces cerevisiae* as the recombination host cell (described in Example 2). Figure 7 shows an overview over the location of the inactivation site of the *Humicola lanuginosa* lipase gene and the number of the clone (referred to as "blue number" in the tables). Location of restriction enzyme sites and clone numbers are relative to the initiation codon of the lipase gene. In all cases a stop codon was located in the new reading frame 10 to 50 bp from the frameshift. Figure 8 shows an overview of the creation of active *humicola lanuginosa* lipase genes from the recombinations in table 2A and B by a "mosaic mechanism". Lines indicate the introduction of the fragment sequence into the vector and lines with a x indicate sequences that are not introduced in the active lipase colonies. The primers used for the PCR fragment are shown together with the location of the frameshift mutation (marked by the restriction site used for the construction). Figure 9 shows an overview of fragments used in the recombination of 2 partial overlapping fragments into a gapped vector. The primers used for the PCR fragments are shown together with the location of the frameshift mutation (if not wild type). Figure 10 shows an overview of fragments used in the recombination of 3 partial overlapping fragments into a gapped vector. The primers used for the PCR fragments are shown. The overlap between PCR353 and 355 is only a 10 bp.

#### DETAILED DESCRIPTION OF THE INVENTION

The object of the present invention is to provide an improved method for preparing positive polypeptide variants by an iterative *in vivo* recombination method.

The inventor of the present invention have surprisingly found an efficient method for shuffling homologous DNA sequences in an *in vivo* recombination system using a eukaryotic cell as a recombination host cell.

A "recombination host cell" is in the context of the present invention a cell capable of mediating shuffling of a number of homologous DNA sequences.

5

The term "shuffling" means recombination of nucleotide sequence(s) between two or more homologous DNA sequences resulting in output DNA sequences (i.e. DNA sequences having been subjected to a shuffling cycle) having a number of nucleotides exchanged, in comparison to 10 the input DNA sequences (i.e. starting point homologous DNA sequences).

An important advantage of the invention is that mosaic DNA sequences with multiple replacement points or replacements, not related to the 15 opening site, is created, which is not discovered in Pompon's method.

An other important advantage of the present invention is that when using a mixture of fragments and opened vectors (in the screening 20 set up) it gives the possibility of many different clones to recombine pairwise or even triplewise (as can be seen in a couple of examples below).

The *in vivo* recombination method of the invention simple to perform 25 and results in a high level of mixing of homologous genes or variants. A large number of variants or homologous genes can be mixed in one transformation. The mixing of improved variants or wild type genes followed by screening increases the number of further improved variants manyfold compared to doing only random 30 mutagenesis.

Recombination of multiple overlapping fragments is possible with a high efficiency increasing the mixing of variants or homologous genes using the *in vivo* recombination method. An overlap as small as 35 10 bp is sufficient for recombination which may be utilized for very easy domain shuffling of even distantly related genes.

The invention relates to a method for preparing polypeptide variants by shuffling different nucleotide sequences of homologous DNA 40 sequences by *in vivo* recombination comprising the steps of

- a) forming at least one circular plasmid comprising a DNA sequence encoding a polypeptide,
- b) opening said circular plasmid(s) within the DNA sequence(s) encoding the polypeptide(s),
- 5 c) preparing at least one DNA fragment comprising a DNA sequence homologous to at least a part of the polypeptide coding region on at least one of the circular plasmid(s), d) introducing at least one of said opened plasmid(s), together with at least one of said homologous DNA fragment(s) covering full-length DNA sequences encoding said polypeptide(s) or parts thereof, into a recombination host cell,
- 10 e) cultivating said recombination host cell, and
- f) screening for positive polypeptide variants.

15 According to the invention more than one cycle of step a) to f) may be performed.

The opening of the plasmid(s) in step b) can be directed toward any site within the polypeptide coding region of the plasmid. The 20 plasmid(s) may be opened by any suitable methods known in the art. The opened ends of the plasmid may be filled-in with nucleotides as described in Pompon et al. (1989), *supra*. It is preferred not to fill in the opened ends as it might create a frameshift.

25 It is preferred to open the plasmid(s) around the middle of the polypeptide coding DNA sequence(s), as this is believed to result in a more effective recombination between DNA fragment(s) and opened plasmid(s).

30 In an embodiment of the invention the DNA fragment(s) is(are) prepared under conditions resulting in a low, medium or high random mutagenesis frequency.

35 To obtain low mutagenesis frequency the DNA sequence(s) (comprising the DNA fragment(s)) may be prepared by a standard PCR amplification method (US 4,683,202 or Saiki et al., (1988), *Science* 239, 487 - 491).

40 A medium or high mutagenesis frequency may be obtained by performing the PCR amplification under conditions which increase the misincorporation of nucleotides, for instance as described by Deshler,

(1992), GATA 9(4), 103-106; Leung et al., (1989), Technique, Vol. 1, No. 1, 11-15.

It is also contemplated according to the invention to combine the 5 PCR amplification (i.e. according to this embodiment also DNA fragment mutation) with a mutagenesis step using a suitable physical or chemical mutagenizing agent, e.g., one which induces transitions, transversions, inversions, scrambling, deletions, and/or insertions.

10 In the context of the present invention the term "positive polypeptide variants" means resulting polypeptide variants possessing functional properties which has been improved in comparison to the polypeptides producible from the corresponding input DNA sequences. Examples, of such improved properties can be as different as e.g. 15 biological activity, enzyme washing performance, antibiotic resistance etc.

Consequently, which screening method to be used for identifying 20 positive variants depend on the desired improved property of the polypeptide variant in question.

If, for instance, the polypeptide in question is an enzyme and the desired improved functional property is the wash performance, the screening in step f) may conveniently be performed by use of a 25 filter assay based on the following principle:

The recombination host cell is incubated on a suitable medium and under suitable conditions for the enzyme to be secreted, the medium being provided with a double filter comprising a first protein-binding filter and on top of that a second filter exhibiting a low 30 protein binding capability. The recombination host cell is located on the second filter. Subsequent to the incubation, the first filter comprising the enzyme secreted from the recombination host cell is separated from the second filter comprising said cells. The first 35 filter is subjected to screening for the desired enzymatic activity and the corresponding microbial colonies present on the second filter are identified.

40 The filter used for binding the enzymatic activity may be any protein binding filter e.g. nylon or nitrocellulose. The topfilter carrying the colonies of the expression organism may be any filter

that has no or low affinity for binding proteins e.g. cellulose acetate or Durapore®. The filter may be pre-treated with any of the conditions to be used for screening or may be treated during the detection of enzymatic activity.

5

The enzymatic activity may be detected by a dye, fluorescence, precipitation, pH indicator, IR-absorbance or any other known technique for detection of enzymatic activity.

10 The detecting compound may be immobilized by any immobilizing agent e.g. agarose, agar, gelatine, polyacrylamide, starch, filter paper, cloth; or any combination of immobilizing agents.

15 If the improved functional property of the polypeptide is not sufficiently good after one cycle of shuffling, the polypeptide may be subjected to another cycle.

20 In an embodiment of the invention at least one shuffling cycle is a backcrossing cycle with the initially used DNA fragment, which may be the wild-type DNA fragment. This eliminates non-essential mutations. Non-essential mutations may also be eliminated by using wild-type DNA fragments as the initially used input DNA material.

25 It is to be understood that the method of the invention is suitable for all types of polypeptide, including enzymes such as proteases, amylases, lipases, cutinases, amylases, cellulases, peroxidases and oxidases.

30 Also contemplated according to the invention is polypeptides having biological activity such as insulin, ACTH, glucagon, somatostatin, somatotropin, thymosin, parathyroid hormone, pigmentary hormones, somatomedin, erythropoietin, luteinizing hormone, chorionic gonadotropin, hypothalamic releasing factors, antidiuretic hormones, thyroid stimulating hormone, relaxin, interferon, thrombopoietin 35 (TPO) and prolactin.

40 Especially contemplated according to the present invention is initially to use input DNA sequences being either wild-type, variant or modified DNA sequences, such as a DNA sequences coding for wild-type, variant or modified enzymes, respectively, in particular enzymes exhibiting lipolytic activity.

In an embodiment of the invention the lipolytic activity is a lipase activity derived from the filamentous fungi of the *Humicola* sp., in particular *Humicola lanuginosa*, especially *Humicola lanuginosa*.

5

In a specific embodiment of the invention the initially used input DNA fragment to be shuffled with a homologous polypeptide is the wild-type DNA sequence encoding the *Humicola lanuginosa* lipase derived from *Humicola lanuginosa* DSM 4109 described in EP 305 216 10 (Novo Nordisk A/S).

Also specifically encompassed by the scope of the invention is input DNA sequences selected from the group of vectors (a) to (f) and/or DNA fragments (g) to (aa) coding for *Humicola lanuginosa* lipase 15 variants from the list below in the Material and Method section.

Throughout the present application the name *Humicola lanuginosa* has been used to identify one preferred parent enzyme, i.e. the one mentioned immediately above. However, in recent years *H. lanuginosa* 20 has also been termed *Thermomyces lanuginosus* (a species introduced the first time by Tsiklinsky in 1989) since the fungus show morphological and physiological similarity to *Thermomyces lanuginosus*. Accordingly, it will be understood that whenever reference is made to *H. lanuginosa* this term could be replaced by 25 *Thermomyces lanuginosus*. The DNA encoding part of the 18S ribosomal gene from *Thermomyces lanuginosus* (or *H. lanuginosa*) have been sequenced. The resulting 18S sequence was compared to other 18S sequences in the GenBank database and a phylogenetic analysis using parsimony (PAUP, Version 3.1.1, Smithsonian Institution, 1993) have 30 also been made. This clearly assigns *Thermomyces lanuginosus* to the class of *Plectomycetes*, probably to the order of *Eurotiales*. According to the Entrez Brower at the NCBI (National Center for Biotechnology Information), this relates *Thermomyces lanuginosus* to families like *Eremascaceae*, *Monoascaceae*, *Pseudoeurotiaceae* and 35 *Trichocomaceae*, the latter containing genera like *Emericella*, *Aspergillus*, *Penicillium*, *Eupenicillium*, *Paecilomyces*, *Talaromyces*, *Thermoascus* and *Sclerocleista*.

Consequently, such genes encoding lipolytic enzymes of filamentous 40 fungi of the genera *Emericella*, *Aspergillus*, *Penicillium*,

*Eupenicillium*, *Paecilomyces*, *Talaromyces*, *Thermoascus* and *Sclerocleista* are also specifically contemplated according to the present invention.

5 Other examples of relevant filamentous fungi genes encoding lipolytic enzymes include strains of the *Absidia* sp. e.g. the strains listed in WO 96/13578 (from Novo Nordisk A/S) which are hereby incorporated by reference. *Absidia* sp. strains listed in WO 96/13578 include *Absidia blakesleeana*, *Absidia corymbifera* and 10 *Absidia reflexa*.

Strains of *Rhizopus* sp., in particular *Rh. niveus* and *Rh. oryzae* are also contemplated according to the invention.

15 The lipolytic gene may also be derived from a bacteria, such as a strain of the *Pseudomonas* sp., in particular *Ps. fragi*, *Ps. stutzeri*, *Ps. cepacia* and *Ps. fluorescens* (WO 89/04361), or *Ps. plantarii* or *Ps. gladioli* (US 4,950,417) or *Ps. alcaligenes* and *Ps. pseudoalcaligenes* (EP 216 272, EP 331 376, or WO 94/25578 20 (disclosing variants of the *Ps. pseudoalcaligenes* lipolytic enzyme), the *Pseudomonas* sp. variants disclosed in EP 407 225, or a *Pseudomonas* sp. lipolytic enzyme, such as the *Ps. mendocina* (also termed *Ps. putida*) lipolytic enzyme described in WO 98/09367 and US 5,389,536 or variants thereof as described in US 5,352,594, or *Ps. 25 euroginosa* or *Ps. glumae*, or *Ps. syringae*, or *Ps. wisconsinensis* (WO 96/12012 from Solvay) or a strain of *Bacillus* sp., e.g. the *B. subtilis* described by Dartois et al., (1993) *Biochimica et Biophysica acta* 1131, 253-260, or *B. stearothermophilus* (JP 64/7744992) or *B. pumilus* (WO 91/16422) or a strain of *Streptomyces* 30 sp., e.g. *S. scabies*, or a strain of *Chromobacterium* sp. e.g. *C. viscosum*.

In connection with the *Pseudomonas* sp. lipases it has been found that lipases from the following organisms have a high degree of 35 homology, such as at least 60% homology, at least 80% homology or at least 90% homology, and thus are contemplated to belong to the same family of lipases: *Ps. ATCC21808*, *Pseudomonas* sp. lipase commercially available as Liposam®, *Ps. aeruginosa* EF2, *Ps. aeruginosa* PAC1R, *Ps. aeruginosa* PAO1, *Ps. aeruginosa* TE 3285, *Ps. 40 sp. 109*, *Ps. pseudoalcaligenes* M1, *Ps. glumae*, *Ps. cepacia* DSM 3959,

5 *Ps. cepacia* M-12-33, *Ps. sp.* KWI-56, *Ps. putida* IFO 3458, *Ps. putida* IFO 12049 (Gilbert, E. J., (1993), *Pseudomonas lipases: Biochemical properties and molecular cloning* Enzyme Microb. Technol., 15, 634-645). The species *Pseudomonas cepacia* has recently been reclassified as *Burkholderia cepacia*, but is termed *Ps. cepacia* in the present application.

10 Also genes encoding lipolytic enzymes from yeasts are relevant, and include lipolytic genes from *Candida* sp., in particular *Candida rugosa*, or *Geotrichum* sp., in particular *Geotrichum candidum*.

15 Specific examples of microorganisms comprising genes encoding lipolytic enzymes used for commercially available products and which may serve as donor of genes to be shuffled according to the invention include *Humicola lanuginosa*, used in Lipolase®, Lipolase® Ultra, *Ps. mendocina* used in Lumafast®, *Ps. alcaligenes* used in Lipomax®, *Fusarium solani*, *Bacillus* sp. (US 5427936, EP 528828), *Ps. mendocina*, used in Liposam®.

20 Also the *Pseudomonas* sp. lipase gene shown in SEQ ID NO 14 are specifically contemplated according to the invention.

25 It is to be emphasized that genes encoding lipolytic enzyme to be shuffled according to the invention may be any of the above mentioned genes of lipolytic enzymes and any variant, modification, or truncation thereof. Examples of such genes which are specifically contemplated include the genes encoding the enzymes described in WO 92/05249, WO 94/01541, WO 94/14951, WO 94/25577, WO 95/22615 and a protein engineered lipase variants as described in EP 407 225; a protein engineered *Ps. mendocina* lipase as described in US 5,352,594; a cutinase variant as described in WO 94/14964; a variant of an *Aspergillus* lipolytic enzyme as described in EP patent 167,309; and *Pseudomonas* sp. lipase described in WO 95/06720.

35 A request to the DNA sequences, encoding the polypeptide(s), to be shuffled, is that they are at least 60%, preferably at least 70%, better more than 80%, especially more than 90%, and even better up to almost 100% homologous. DNA sequences being less homologous will have less inclination to interact and recombine.

40 It is also contemplated according to the invention to shuffle parent (homologous) wild type organisms of different genera.

Further, the DNA fragment(s) to be shuffled may preferably have a length of from about 20 bp to 8 kb, preferably about 40 bp to 6 kb, more preferred about 80 bp to 4 kb, especially about 100 bp to 2 kb,  
5 to be able to interact optimally with the opened plasmid.

The method of the invention is very efficient for preparing polypeptide variants in comparison to prior art method comprising transforming linear DNA fragments/sequences.

10 The inventor found that the transformation frequency of a mixture of opened plasmid and a DNA fragment were significantly higher than when transforming a plasmid cut at the same site alone. The transformation frequency of the opened plasmid and DNA fragment were as  
15 high as for uncut plasmid.

Without being limited to any theory it is believed that the opening of the plasmid(s) restrict(s) the replication of (opened) plasmid(s) when not interacting with at least one DNA fragment. In accordance  
20 with this an increased number of recombined DNA sequences were found after only one shuffling cycle.

As described in Example 1 50% of the resulting transformants contained recombined DNA sequences of both input DNA sequences. As  
25 high as 20% of the total number of recombined DNA sequences were "random" mixtures (i.e. having more than one region of nucleotides exchanged).

30 The input DNA sequences may be any DNA sequences including wild-type DNA sequences, DNA sequences encoding variants or mutants, or modifications thereof, such as extended or elongated DNA sequences, and may also be the outcome of DNA sequences having been subjected to one or more cycles of shuffling (i.e. output DNA sequences) according to the method of the invention or any other method (e.g.  
35 any of the methods described in the prior art section).

When using the method of the invention the output DNA sequences (i.e. shuffled DNA sequences), have had a number of nucleotide(s) exchanged. This results in replacement of at least one amino acid within the polypeptide variant, if comparing it with the parent polypeptide. It is to be understood that also silent mutations is  
40

contemplated (i.e. nucleotide exchange which does not result in changes in the amino acid sequence).

However, the method of the present invention will in most cases lead 5 to the replacement of a considerable number of amino acid and may in certain cases even alter the structure of one or more polypeptide domains (i.e. a folded unit of polypeptide structure).

According to the present invention more than two DNA sequences are 10 shuffled at the same time. Actually any number of different DNA fragments and homologous polypeptides comprised in suitable plasmids may be shuffled at the same time. This is advantageous as a vast number of quite different variants can be made rapidly without an abundance of iterative procedures.

15 The inventor have tested the nucleotide shuffling method of the invention using significantly more than two homologous DNA sequences. As described in Example 2 it was surprisingly found that the method of the invention advantageously can be used for 20 recombining more than two DNA sequences.

One cycle of shuffling according to the method of the invention may result in the exchange of from 1 to 1000 nucleotides into the opened 25 plasmid DNA sequence encoding the polypeptide in question. The exchanged nucleotide sequence(s) may be continuous or may be present as a number of sub-sequences within the full-length sequence(s).

To support the present invention the inventor made a number of additional experiments on different aspect on the method of the 30 invention. The experiments are described below and illustrated in the Example 3 to 6 below.

A number of vectors and fragments comprising an inactivated synthetic *Humicola lanuginosa* lipase genes were constructed by 35 introducing frameshift/stop codon mutations in the lipase gene at various positions. These were used for monitoring the *in vivo* recombination of different combinations of opened vector(s) and DNA fragments. The number of active lipase colonies were scored as described in Example 3. The number of colonies determines the 40 efficiency of the opened vector(s) and fragment(s) recombination.

One frameshift mutation in said *Humicola lanuginosa* lipase gene in the opened vector and another in the fragment on the opposite side of the opening site gave 3 to 32% of active lipase colonies depending on the location and combination. It was concluded that

5 the closer that the mutation is at the ends of the vector the higher mixing.

One frameshift mutation in the opened vector and two in the fragment on each side of the opening site gave 4 to 42% of active

10 colonies depending on the location and combination. Some of these active colonies can be considered to be mosaics, not only related to the opening site.

Two frameshift mutations in the opened vector on each side of the

15 opening site and one in the fragment gave 0.5 to 3.1% of active colonies depending on the location and combination. Most of these active colonies are mosaics of the "parent" DNA.

Two frameshift mutations in the opened vector on each side of the

20 opening site and a wild type fragment gave 7.7 to 10.7% of active colonies depending on the location.

It was also found that the amount of vectors relative to fragments and the size of the fragments are also influencing the result.

25 Using of the *S. cerevisiae* rad52 mutants as the recombination host cell showed that the rad52 mutant transformed very well with wild type plasmid(s) and expressed the *Humicola lanuginosa* lipase gene, but gave no transformants at all with the opened vectors and

30 fragments.

The RAD52 function is required for "classical recombination" (but not for unequal sister-strand mitotic recombination) showing that the recombination of opened vector and fragment could involve a

35 classical recombination mechanism.

Classical recombination is the recombination mechanism involved in the recombination between genes located on nonsister chromatids of homologous chromosomes as defined in for example Petes TD, Malone

40 RE and Symington LS (1991) "Recombination in Yeast", page 407-522, in The Molecular and Cellular Biology of the Yeast *Saccharomyces*,

Volume 1 (eds. Broach JR, Pringle JR and Jones EW), Cold Spring Harbor Laboratory Press, New York.

Multiple partially overlapping fragments

5 The inventor also tested recombination of multiple partial overlapping fragments using the method of the invention.

10 The recombination of 2 and 3 partial overlapping fragments into a gapped (i.e. that the opening result in cutting out of a little part of the gene) vector were tested and gave a high recovery of recombined *Humicola lanuginosa* lipase gene. The recovery of active lipase gene from different combinations of inactivated *Humicola lanuginosa* genes was tested for the recombination of 2 partial overlapping fragments. The tendency was a higher mixing in the 15 overlapping region between the 2 fragments in the gapped region than in the vector and fragment overlap.

20 When recombining many fragments from the same region, the multiple overlapping fragment technique will increase the mixing by itself, but it is also important to have a relative high random mixing in overlapping regions in order to mix closely located variants/differences.

25 An overlap as small as 10 bp between two fragments were found to be sufficient to obtain a very efficient recombination. Therefore, overlapping in the range from 5 to 5000 bp, preferably from 10 bp to 500 bp, especially 10 bp to 100 bp is suitable according to the method of the invention.

30 According to this embodiment of the present invention 2 or more overlapping fragments, preferable 2 to 6 overlapping fragments, especially 2 to 4 overlapping fragments may advantageously be used as input fragments in a shuffling cycle.

35 Besides increasing the mixing of genes, this is a very useful method for domain shuffling by creating small overlaps between DNA fragments from different domains and screen for the best combination.

40 For instance, in the case of three DNA fragments the overlapping regions may be as follows:

- the first end of the first fragment overlaps the first end of the opened plasmid,
- the first end of the second fragment overlaps the second end of the first fragment, and the second end of the second fragment overlaps the first end of the third fragment,
- 5 - the first end of the third fragment overlaps (as stated above) the second end of the second fragment, and the second end of the third fragment overlaps the second end of the opened plasmid.

10 It is to be understood that when using two or more DNA fragments as starting material it is preferred to have continuos overlaps between the ends of the plasmid and the DNA-fragments.

15 Even though it is preferred to shuffle homologous DNA sequences in the form of DNA fragment(s) and opened plasmid(s), it is also contemplated according to the invention to shuffle two or more opened plasmids comprising homologous DNA sequences encoding polypeptides. However, in such case it is compulsory to open the plasmids at different sites.

20 In an further embodiment of the invention two or more opened plasmids and one or more homologous DNA fragments are used as the starting material to be shuffled. The ratio between the opened plasmid(s) and homologous DNA fragment(s) preferably lie in the range from 20:1 to 1:50, preferable from 2:1 to 1:10 (mol vector: mol fragments) with the specific concentrations being from 1 pM to 10 M of the DNA.

25 The opened plasmids may advantagously be gapped in such a way that the overlap between the fragments is deleted in the vector in order to select for the recombination).

Preparing the DNA fragment

The DNA fragment to be shuffled with the homologous polypeptide comprised in an opened plasmid may be prepared by any suitable method. For instance, the DNA fragment may be prepared by PCR amplification (polymerase chain reaction), as described above, of a plasmid or vector comprising the gene of the polypeptide, using specific primers, for instance as described in US 4,683,202 or Saiki et al., (1988), Science 239, 487 - 491. The DNA fragment may also be cut out from a vector or plasmid comprising the desired DNA sequence by digestion with restriction enzymes, followed by isolation using e.g. electrophoresis.

The DNA fragment encoding the homologous polypeptide in question may alternatively be prepared synthetically by established standard methods, e.g. the phosphoamidite method described by Beaucage and Caruthers, (1981), Tetrahedron Letters 22, 1859 - 1869, or the method described by Matthes et al., (1984), EMBO Journal 3, 801 - 805. According to the phosphoamidite method, oligonucleotides are synthesized, e.g. in an automatic DNA synthesizer, purified, annealed, ligated and cloned in suitable vectors.

Furthermore, the DNA fragment may be of mixed synthetic and genomic, mixed synthetic and cDNA or mixed genomic and cDNA origin prepared by ligating fragments of synthetic, genomic or cDNA origin (as appropriate), the fragments corresponding to various parts of the entire DNA sequence, in accordance with standard techniques.

The plasmid

The plasmid comprising the DNA sequence encoding the polypeptide in question may be prepared by ligating said DNA sequence into a suitable vector or plasmid, or by any other suitable method.

Said vector may be any vector which may conveniently be subjected to recombinant DNA procedures. The choice of vector will often depend on the recombination host cell into which it is to be introduced.

Thus, the vector may be an autonomously replicating vector, i.e. a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g. a plasmid. Alternatively, the vector may be one which, when introduced into the recombination host cell, is integrated into the host cell genome and

replicated together with the chromosome(s) into which it has been integrated.

To facilitate the screening process it is preferred that the vector 5 is an expression vector in which the DNA sequence encoding the polypeptide in question is operably linked to additional segments required for transcription of the DNA. In general, the expression vector is derived from a plasmid, a cosmid or a bacteriophage, or may contain elements of any or all of these.

10

The term, "operably linked" indicates that the segments are arranged so that they function in concert for their intended purposes, e.g. transcription initiates in a promoter and proceeds through the DNA sequence coding for the polypeptide in question.

15

The promoter may be any DNA sequence which shows transcriptional activity in the recombination host cell of choice and may be derived from genes encoding proteins, such as enzymes, either homologous or heterologous to the host cell.

20

Examples of suitable promoters for use in yeast host cells include promoters from yeast glycolytic genes (Hitzeman et al., (1980), J. Biol. Chem. 255, 12073 - 12080; Alber and Kawasaki, (1982), J. Mol. Appl. Gen. 1, 419 - 434) or alcohol dehydrogenase genes (Young et 25 al., in Genetic Engineering of Microorganisms for Chemicals (Hollaender et al, eds.), Plenum Press, New York, 1982), or the TPI1 (US 4,599,311) or ADH2-4c (Russell et al., (1983), Nature 304, 652 - 654) promoters.

30 Examples of suitable promoters for use in filamentous fungus host cells are, for instance, the ADH3 promoter (McKnight et al., (1985), The EMBO J. 4, 2093 - 2099) or the tpiA promoter. Examples of other useful promoters are those derived from the gene encoding *A. oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *A. niger* neutral  $\alpha$ -amylase, *A. niger* acid stable  $\alpha$ -amylase, *A. niger* or *A. awamori* glucoamylase (gluA), *Rhizomucor miehei* lipase, *A. oryzae* alkaline protease, *A. oryzae* triose phosphate isomerase or *A. nidulans* acetamidase. Preferred are the TAKA-amylase and gluA 35 promoters.

40

The DNA sequence encoding polypeptide in question invention may also, if necessary, be operably connected to a suitable terminator, such as the human growth hormone terminator (Palmiter et al., op. cit.) or (for fungal hosts) the TPII (Alber and Kawasaki, op. cit.) 5 or ADH3 (McKnight et al., op. cit.) terminators. The vector may further comprise elements such as polyadenylation signals (e.g. from SV40 or the adenovirus 5 Elb region), transcriptional enhancer sequences (e.g. the SV40 enhancer) and translational enhancer sequences (e.g. the ones encoding adenovirus VA RNAs).

10

The vector may further comprise a DNA sequence enabling the vector to replicate in the recombination host cell in question.

When the host cell is a yeast cell, suitable sequences enabling the vector to replicate are the yeast plasmid 2m replication genes REP 15 1-3 and origin of replication.

The plasmid pY1 can be used for production of useful proteins and peptides, using filamentous fungi, such as *Aspergillus* sp., and yeasts as recombinant host cells (JP06245777-A).

20

The vector may also comprise a selectable marker, e.g. a gene the product of which complements a defect in the recombination host cell, such as the gene coding for dihydrofolate reductase (DHFR) or the *Schizosaccharomyces pombe* TPI gene (described by P.R. Russell, 25 (1985), Gene 40, 125-130).

Another example of such suitable selective markers are the *ura3* and *leu2* genes which complements the corresponding defect genes of e.g. the yeast strain *Saccharomyces cerevisiae* YNG318.

30

The vector may also comprise a selectable marker which confers resistance to a drug, e.g. ampicillin, kanamycin, tetracycline, chloramphenicol, neomycin, hygromycin or methotrexate. For filamentous fungi, selectable markers include amdS, pyrG, argB, niaD, 35 sC, trpC, pyr4, and DHFR.

To direct the polypeptide in question into the secretory pathway of the recombination host cell, a secretory signal sequence (also known as a leader sequence, prepro sequence or pre sequence) may be 40 provided in the recombinant vector. The secretory signal sequence is joined to the DNA sequence encoding the lipolytic enzyme in the

correct reading frame. Secretory signal sequences are commonly positioned 5' to the DNA sequence encoding the polypeptide. The secretory signal sequence may be the signal normally associated with the polypeptide in question or may be from a gene encoding another 5 secreted protein.

The signal peptide may be naturally occurring signal peptide, or a functional part thereof, or it may be a synthetic peptide. For secretion from yeast cells, suitable signal peptides have been found 10 to be the  $\alpha$ -factor signal peptide (cf. US 4,870,008), the signal peptide of mouse salivary amylase (cf. O. Hagenbuchle et al., (1981), Nature 289, 643-646), a modified carboxypeptidase signal peptide (cf. L.A. Valls et al., (1987), Cell 48, 887-897), the *Humicola lanuginosa* lipase signal peptide, the yeast BAR1 signal 15 peptide (cf. WO 87/02670), or the yeast aspartic protease 3 (YAP3) signal peptide (cf. M. Egel-Mitani et al., (1990), Yeast 6, 127-137).

For efficient secretion in yeast, a sequence encoding a leader 20 peptide may also be inserted downstream of the signal sequence and upstream of the DNA sequence encoding the polypeptide in question. The function of the leader peptide is to allow the expressed polypeptide to be directed from the endoplasmic reticulum to the Golgi apparatus and further to a secretory vesicle for secretion 25 into the culture medium (i.e. exportation of the polypeptide across the cell wall or at least through the cellular membrane into the periplasmic space of the yeast cell). The leader peptide may be the yeast  $\alpha$ -factor leader (the use of which is described in e.g. US 4,546,082, EP 16 201, EP 123 294, EP 123 544 and EP 163 529). 30 Alternatively, the leader peptide may be a synthetic leader peptide, which is to say a leader peptide not found in nature. Synthetic leader peptides may, for instance, be constructed as described in WO 89/02463 or WO 92/11378.

35 For use in filamentous fungi, the signal peptide may conveniently be derived from a gene encoding an *Aspergillus* sp. amylase or glucoamylase, a gene encoding a *Rhizomucor miehei* lipase or protease, a *Humicola lanuginosa* lipase. The signal peptide is preferably derived from a gene encoding *A. oryzae* TAKA amylase, *A. 40 niger* neutral  $\alpha$ -amylase, *A. niger* acid-stable amylase, or *A. niger*

glucoamylase.

The recombination host cell

5 The recombination host cell, into which the mixture of plasmid/fragment DNA sequences are to be introduced, may be any eukaryotic cell, including fungal cells and plant cells, capable of recombining the homologous DNA sequences in question.

10 According to prior art prokaryotic microorganisms, such as bacteria including *Bacillus* and *E. coli*; eukaryotic organisms, such as filamentous fungi, including *Aspergillus* and yeasts such as *Saccharomyces cerevisiae*; and tissue culture cells from avian or mammalian origins have been suggested for *in vivo* recombination. All 15 of said organisms can be used as recombination host cell, but in general prokaryotic cells are not sufficiently effective (i.e. does not result in a sufficient number of variants) to be suitable for recombination methods for industrial use.

20 Consequently, preferred recombination host cells according to the present invention are fungal cells, such as yeast cells or filamentous fungi.

25 Examples of suitable yeast cells include cells of *Saccharomyces* sp., in particular strains of *Saccharomyces cerevisiae* or *Saccharomyces kluyveri* or *Schizosaccharomyces* sp.. Methods for transforming yeast cells with heterologous DNA and producing heterologous polypeptides therefrom are described, e.g. in US 4,599,311, US 4,931,373, US 4,870,008, 5,037,743, and US 4,845,075, all of which are hereby incorporated by reference. Transformed cells may be selected by, 30 e.g., a phenotype determined by a selectable marker, commonly drug resistance or the ability to grow in the absence of a particular nutrient, e.g. leucine. A preferred vector for use in yeast is the POT1 vector disclosed in US 4,931,373. The DNA sequence encoding the polypeptide may be preceded by a signal sequence and optionally a 35 leader sequence, e.g. as described above. Further examples of suitable yeast cells are strains of *Kluyveromyces*, such as *K. lactis*, *Hansenula*, e.g. *H. polymorpha*, or *Pichia*, e.g. *P. pastoris* (cf Gleeson et al., (1986), J. Gen. Microbiol. 132, 3459-3465; US 4,882,279).

Examples of other fungal cells are cells of filamentous fungi, e.g. *Aspergillus* sp., *Neurospora* sp., *Fusarium* sp. or *Trichoderma* sp., in particular strains of *A. oryzae*, *A. nidulans* or *A. niger*. The use of *Aspergillus* sp. for the expression of proteins is described in, 5 e.g., EP 272 277, EP 230 023. The transformation of *F. oxysporum* may, for instance, be carried out as described by Malardier et al., (1989), Gene 78, 147-156.

10 In a preferred embodiment of the invention the recombination host cell is a cell of the genus *Saccharomyces*, in particular *S. cerevisiae*.

#### METHODS AND MATERIALS

15

##### DNA sequence:

*Humicola lanuginosa* DSM 4109 derived lipase encoding DNA sequence.

*Humicola lanuginosa* lipase variants:

20 Variants used for preparing vectors to be opened with NruI in Example 2:

- (a) E56R, D57L, I90F, D96L, E99K
- (b) E56R, D57L, V60M, D62N, S83T, D96P, D102E
- (c) D57G, N94K, D96L, L97M
- 25 (d) E87K, G91A, D96R, I100V, E129K, K237M, I252L, P256T, G263A, L264Q
- (e) E56R, D57G, S58F, D62C, T64R, E87G, G91A, F95L, D96P, K98I, (K237M)
- (f) E210K

30 Variants used for preparing DNA fragments by standard PCR amplification in Example 2:

- (g) S83T, N94K, D96N
- (h) E87K, D96V
- (i) N94K, D96A
- (j) E87K, G91A, D96A
- 35 (k) D167G, E210V
- (l) S83T, G91A, Q249R
- (m) E87K, G91A
- (n) S83T, E87K, G91A, N94K, D96N, D111N.
- (o) N73D, E87K, G91A, N94I, D96G.
- 40 (p) L67P, I76V, S83T, E87N, I90N, G91A, D96A, K98R.

(q) S83T, E87K, G91A, N92H, N94K, D96M  
(s) S85P, E87K, G91A, D96L, L97V.  
(t) E87K, I90N, G91A, N94S, D96N, I100T.  
(u) I34V, S54P, F80L, S85T, D96G, R108W, G109V, D111G, S116P, L124S,  
5 V132M, V140Q, V141A, F142S, H145R, N162T, I166V, F181P, F183S,  
R205G, A243T, D254G, F262L.  
(v) E56R, D57L, I90F, D96L, E99K  
(x) E56R, D57L, V60M, D62N, S83T, D96P, D102E  
(y) D57G, N94K, D96L, L97M  
10 (z) E87K, G91A, D96R, I100V, E129K, K237M, I252L, P256T, G263A, L264Q  
(aa) E56R, D57G, S58F, D62C, T64R, E87G, G91A, F95L, D96P, K98I

Strains:

Expression system host:

15 *Saccharomyces cerevisiae* YNG318: MAT $\alpha$  Dpep4<sup>[cir]</sup> ura3-52, leu2-D2,  
his 4-539  
*Saccharomyces cerevisiae* Rad52: Strain M1533 = MAT $\alpha$  rad52 ura3,  
obtained from Torsten Nilsson Tillgren, Institute of Genetics,  
University of Copenhagen.

20

Plasmids:

pJS026 (see figure 3)  
pJS037 (see figure 4)  
pYES 2.0 (Invitrogen)

25

Transformation selective marker

ura3

leu2

30 Media

SC-ura<sup>-</sup>: 90 ml 10 x Basal salt, 22.5 ml 20% casamino acids, 9 ml 1%  
tryptophan, H<sub>2</sub>O ad 806 ml, autoclaved, 3.6 ml 5% threonine and 90 ml  
20% glucose or 20% galactose added.

LB-medium: 10 g Bacto-tryptone, 5 g Bacto yeast extract, 10 g NaCl  
35 in 1 litre water.

Brilliant Green (BG) (Merck, art. No. 1.01310)

BG-reagent: 4 mg/ml Brilliant Green (BG) dissolved in water

Substrate 1:

10 ml olive oil (Sigma CAT NO. 0-1500)

40 20 ml 2% polyvinyl alcohol (PVA)

The Substrate is homogenised for 15-20 minutes.

Methods:

5 Construction of yeast expression vector

The expression plasmids pJS026 and pJS037, are derived from pYES 2.0. The inducible GAL1-promoter of pYES 2.0 was replaced with the constitutively expressed TPI (triose phosphate isomerase)-promoter from *Saccharomyces cerevisiae* (Albert and Kärwasaki, (1982), J. Mol. Appl. Genet., 1, 419-434), and the ura3 promoter has been deleted. A restriction map of pJS026 and pJS037 is shown in figure 3 and figure 4, respectively.

Preparation of the wild-type DNA fragment

15 A lipase wild-type DNA fragment can be prepared either by PCR amplification (resulting in low, medium or high mutagenesis), of the pJS026 plasmid or by cutting the DNA fragment out by digesting with a suitable restriction enzyme.

20 Fermentation of *Humicola lanuginosa* lipase variants in yeast

10 ml of SC-ura<sup>-</sup> medium is inoculated with a *S. cerevisiae* colony and grown at 30°C for 2 days. The 10 ml is used for inoculating 300 ml SC-ura<sup>-</sup> medium which is grown at 30°C for 3 days. The 300 ml is used for inoculation 5 l of the following G-substrate:

25 400 g Amicase

6.7 g yeast extract (Difco)

12.5 g L-Leucin (Fluka)

6.7 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>

10 g MgSO<sub>4</sub>·7H<sub>2</sub>O

30 17 g K<sub>2</sub>SO<sub>4</sub>

10 ml Trace compounds

5 ml Vitamin solution

6.7 ml H<sub>3</sub>PO<sub>4</sub>

25 ml 20% Pluronic (antifoam)

35

In a total volume of 5000 ml:

The yeast cells are fermented for 5 days at 30°C. They are given a start dosage of 100 ml 70% glucose and added 400 ml 70% glucose/day. A pH=5.0 is kept by addition of a 10% NH<sub>3</sub> solution. Agitation is 300 rpm for the first 22 hours followed by 900 rpm for the rest of the fermentation. Air is given with 1 l air/l/min for the first 22 hours

followed by 1.5 l air/l/min for the rest of the fermentation.

Trace compounds:

6.8 g ZnCl<sub>2</sub>  
5 54.0 g FeCl<sub>2</sub>·6H<sub>2</sub>O  
19.1 g MnCl<sub>2</sub>·4H<sub>2</sub>O  
2.2 g CuSO<sub>4</sub>·5H<sub>2</sub>O  
2.58 g CoCl<sub>2</sub>  
0.62 g H<sub>3</sub>BO<sub>3</sub>  
10 0.024 g (NH<sub>4</sub>)<sub>6</sub>Mo<sub>7</sub>O<sub>24</sub>·4H<sub>2</sub>O  
0.2 g KI  
100 ml HCl (concentrated)  
In a total volume of 1 l.

15 Vitamin solution:

250 mg Biotin  
3 g Thiamin  
10 g D-Calciumpanthetonat  
100 g Myo-Inositol  
20 50 g Cholinchlorid  
1.6 g Pyridoxin  
1.2 g Niacinamid  
0.4 g Folicacid  
0.4 g Riboflavin  
25 In a total volume of 1 l.

Transformation of yeast

*Saccharomyces cerevisiae* is transformed by standard methods (cf. Sambrooks et al., (1989), Molecular Cloning: A Laboratory Manual, 30 2nd Ed., Cold Spring Harbor)

Determination of yeast transformation frequency

The transformation frequency is determined by cultivating the transformants on SC-ura<sup>+</sup>plates for 3 days and counting the number of 35 colonies appearing. The number of transformants per mg opened plasmid is the transformation frequency.

Screening for positive variants with improved wash performance

The following filter assay can be used for screening positive 40 variants with improved wash performance.

Low calcium filter assay

- 1) Provide SC Ura<sup>r</sup> replica plates (useful for selecting strains carrying the expression vector) with a first protein binding filter (Nylon membrane) and a second low protein binding filter (Cellulose acetate) on the top.
- 5 2) Spread yeast cells containing a parent lipase gene or a mutated lipase gene on the double filter and incubate for 2 or 3 days at 30°C.
- 10 3) Keep the colonies on the top filter by transferring the top-filter to a new plate.
- 4) Remove the protein binding filter to an empty petri dish.
- 5) Pour an agarose solution comprising an olive oil emulsion (2% PVA:olive oil=3:1), Brilliant green (indicator,0.004%), 100 mM tris buffer pH9 and EGTA (final concentration 5mM) on the bottom filter so as to identify colonies expressing lipase activity in the form of blue-green spots.
- 15 6) Identify colonies found in step 5) having a reduced dependency for calcium as compared to the parent lipase.

20

DNA sequencing was performed by using applied Biosystems ABI DNA sequence model 373A according to the protocol in the ABI Dye Terminator Cycle Sequencing kit.

25 Assessing the efficiency of recombination

The number of colonies determines the efficiency of the opened vector and fragment recombination. The percentage of colonies with active lipase activity gives an estimate of the mixing of the active and inactive genes - theoretically it can be calculated for 30 one frameshift that the closer to 50% the better mixing if equal likelihood of wild type and frameshift, 25% for 2 frameshifts and 12.5% for 3 frameshifts.

Frameshift mutation .

- 35 The frameshift mutation were created either by filling in a restriction site (in case of 5' overhang) or deleting the "sticky ends" (in case of 3' overhang) by T4 DNA polymerase with or without dNTP (deoxynucleotides = equal amounts of dATP, dTTP, dCTP and dGTP). Methods for filling in of restriction sites (referred to as 40 "F" on Figure 7) and deleting the sticky ends (referred to as "(D)" on Figure 7) are well known in the art.

Method for assessing colonies with lipase activity

The number of colonies and positives (i.e. with lipase activity) are calculated as the average of 3 plates.

- 5 The cultivation condition and screening condition used is the following:
  - 1) Provide SC Ura-plates with a protein binding filter (Nylon filter) onto the plate.
  - 2) Spread yeast cells containing a parent lipase gene or a mutated lipase gene on the filter and incubate for 3 or 4 days at 30°C.
  - 3) Remove the protein binding filter with the colonies to a petri dish containing: An agarose solution comprising an olive oil emulsion (2% PVA:Olive oil=2:1), Brilliant green (indicator, 0.004%), 100 mM tris buffer pH 9.
  - 15 5) Identify colonies expressing lipase activity in the form of blue-green spots.

EXAMPLES

20

Example 1Testing in vivo recombination of two homologous genes

The *Saccharomyces cerevisiae* expression plasmid pJS026 was constructed as described above in the "Material and Methods"-section.

A synthetic *Humicola lanuginosa* lipase gene (in pJS037) containing 12 additional restriction sites (see figure 4) was cut with NruI, 30 PstI, and NruI and PstI, respectively, to open the gene approximately in the middle of the DNA sequence encoding the lipase.

The opened plasmid (pJS037) was transformed into *Saccharomyces cerevisiae* YNG318 together with an about 0.9 kb wild-type *Humicola lanuginosa* lipase DNA fragment (see figure 1) prepared from pJS026 by PCR amplification.

Further, the opened plasmid was also transformed into the yeast recombination host cell alone (i.e. without the 0.9 kb synthetic 40 lipase DNA fragment).

The transformed yeast cells were grown as described in the "Ma-

terials and Method"-section above, and the transformation frequency was determined as described above.

It was found that the transformation frequency of the opened plasmid 5 alone was very low (10 transformants per mg opened plasmid), in comparison to the transformation frequency of said plasmid/fragment (50,000 transformants per mg opened plasmid).

The plasmid/fragment was PCR amplified resulting in 20 transformants 10 containing fragments covering the lipase gene region of the recombined plasmid/fragments. The recombination mixture of the 20 transformants were analyzed by restriction site digestion using standard methods. The result is displayed in Table 1.

15 Table 1

20	PCR fragment	NruI (not tested)							
		SphI HindIII PstI			BstXI NheI BstEII KpnI XbaI				
		wt	wt	wt	wt	wt	wt	wt	wt
P1	wt	wt	wt	wt	wt	wt	wt	wt	wt
P2	sg	sg	sg	wt	wt	wt	wt	wt	wt
P3	sg	sg	sg	sg	nd	sg	sg	nd	
P4	nd	sg	sg	wt	nd	wt	nd	nd	
P5	wt	wt	nd	wt	wt	wt	wt	wt	
P6	sg	sg	sg	sg	sg	sg	sg	nd	
N1	wt	wt	wt	wt	sg	wt	wt	wt	
N2	wt	wt	wt	wt	wt	wt	wt	wt	
30	N3	wt	wt	wt	wt	wt	wt	wt	
N4	sg	sg	sg	wt	wt	wt	wt	wt	
N5	sg	sg	sg	wt	wt	wt	wt	wt	
N6	wt	wt	wt	sg	sg	sg	sg	sg	
35	P/N1	sg	sg	sg	wt	wt	wt	wt	wt
P/N2	sg	sg	sg	sg	sg	sg	sg	nd	
P/N3	sg	sg	sg	wt	nd	sg	sg	sg	
P/N4	sg	sg	sg	sg	sg	sg	sg	nd	
P/N5	sg	sg	sg	sg	sg	sg	sg	nd	
40	P/N6	sg	sg	sg	wt	nd	sg	sg	sg
P/N7	nd	wt	wt	wt	nd	wt	nd	wt	
P/N8	sg	sg	sg	wt	wt	wt	sg	nd	

P: plasmid opened with PstI

45 N: Plasmid opened with NruI

P/N: plasmid opened with PstI and NruI (resulting in the removal of a 75 bp fragment)

wt: wild-type gene restriction enzyme pattern

sg: synthetic gene restriction enzyme pattern

50 nd: not determined

As can be seen from Table 1 10 transformants (equivalent to 50%)

contained recombined DNA sequences. 4 of these 10 DNA sequences (equivalent to 20%) contained either a region of the wild-type gene recombined into the synthetic gene or a region of the synthetic gene recombined into the wild-type fragment.

5

Example 2

In vivo recombination of *Humicola lanuginosa* lipase variants

10 The DNA sequences of 20 variants of the *Humicola lanuginosa* lipase were in vivo recombined in the same mixture.

Six vectors were prepared from the lipase variants (a) to (f) (see the list above) by ligation into the yeast expression vector pJS037. All vectors were cut open with *Nru*1.

15

DNA fragment of all 20 homologous DNA sequences (g) to (aa) (see the list above) were prepared by PCR amplification using standard methods.

20 The 20 DNA fragments and the 6 opened vectors were mixed and transformed into the yeast *Saccharomyces cerevisiae* YNG318 by standard methods. The recombination host cell was cultivated as described above and screened as described above. About 20 transformants were isolated and tested for improved wash performance  
25 using the filter assay method described in the "Material and Methods"-section.

Two positive transformants (named A and B) were identified using the filter assay.

30

In comparison to the wild-type amino acid sequence the two recombined positive transformants had the following mutations.

35

A: D57G, N94K, D96L, P256T  
---- ---- ---- =====

A is a recombination of two variants.

---- originates from the vector (d)

===== originates from the DNA fragment prepared from variant (y)

40

B: D57G, G59V, N94K, D96L, L97M, S116P, S170P, N249R  
---- ??? ---- ---- <<<< ??? ==

B is a recombination of vector (c), DNA fragments (n) and (u).

----- originates from the vector (c)  
<<< originates from the DNA fragment prepared from variant (u)  
===== originates from the DNA fragment prepared from variant (n)  
???? Amino acid mutation which is not a result of recombination.

5

As can be seen the resulting positive variants have been formed by recombination two or more variants. The amino acid mutations marked "?????" are not a result of *in vivo* recombination, as none of the shuffled lipase variants (see the list above) comprise any of said 10 mutations. Consequently, these mutations are a result of random mutagenesis arisen during preparation of the DNA fragments by standard PCR amplification.

15 Example 3

Recombination with one frameshift mutations

Synthetic *Humicola lanuginosa* lipase gene (in vector JS037) was 20 made inactive at various positions by deleting (positions 184/385) or filling-in (position 290/317/518/746) restriction enzyme sites or by site-directed introduction of a stop codon. All inactive synthetic lipase genes of 900 bp can be deduced from Figure 7).

25 A number of different 900 bp DNA fragments were made from the above vectors using primer 4699 and primer 5164 using standard PCR technique. Smaller PCR fragments were made using primer 8487 and primer 4548 (260bp), primer 2843 and primer 4548 (488bp).

30 0.5 ml (app. 0.1 mg) of vectors Blue 425, Blue 426, Blue 428 and Blue 429, opened with Pst I (i.e. position 385), vectors Blue 424 and Blue 425 opened with NruI (i.e. position 464) were together with 3 ml (app. 0.5 mg) of fragments 424, 425, 426, 428, 429 in various combination transformed into 100 ml *Saccharomyces cerevisiae* 35 YNG318 competent cells as displayed in Table 1A.

40 The number of colonies and positives (i.e. with lipase activity) were calculated as the average of 3 plates as described in the Material and Methods section.

40

The result of the test is shown in Table 1A

Table 1A

vector + Fragment	Number of colonies	% of colonies with active lipase activity
1. Blue 428 + 429 <sup>a</sup>	774	16%
2. Blue 429 + 428 <sup>#</sup>	645	3%
3. Blue 426 + 425 <sup>#</sup>	276	25%
4. Blue 425 + 426	528	18%
5. Blue 425/Nru I + 426	539	28%
6. Blue 425 + 424	139	7%
7. Blue 424/NruI + 425 <sup>a</sup>	74	32%
8. Blue 428 + 425	81	12%
9. Blue 428 + wt fragment	317	37%

5 Pairwise recombinations of one frameshift mutation on the vector and another on the fragment on the opposite side of the opening site. <sup>a</sup> determined by 9 plates; <sup>#</sup> determined by 6 plates.

10 The first 2 rows of Table 1A displays vectors and fragments with a frameshift on each side of the PstI site. The "mirror image" experiment in row 2 compared to row 1 gives a reproducible lower number of active colonies. The same is true for row 3 and 4 even though it is not as pronounced. Moving the opening site closer to the frameshift in the vector increases the number of actives as seen in row 5. This can explain the reason for the difference in the "mirror image" experiments. In both cases the higher number of 15 positives has the opening site closer to the frameshift in the vector.

20 It can therefore be concluded that the closer the mutation is to the end of the vector the higher chance of mixing. This is probably arising from the well known fact that free DNA ends have a high recombinogenic potential. Therefore it is desirable to have as many free DNA ends as possible to increase the mixing of the genes. This is for example obtained in the later example with recombination of multiple overlapping fragments.

25

Row 6 has a rather low number of actives probably due to the location of the frameshift on the fragment exactly at the PstI opening site of the vector.

Row 7 has the frameshift of the vector close to the opening site and again it gives a high number of actives.

Recombination with one stop codon mutations

5 In order to test if there are any difference in the recombination efficiency of stop codon mutations compared to frameshift mutations the following experiments were made..

10 The same way as described above 0.5 ml (app. 0.1 mg) vectors Blue 624, Blue 625 and Blue 626 (see Table 1B) opened with PstI comprising stop codons at specified positions (positions 184, 317 and 746, respectively) (prepared by site-directed mutagenesis) were together with 3 ml (app. 0.5 mg) of fragments 624, 625 and 626 transformed into 100 ml *Saccharomyces cerevisiae* YNG318 competent

15 cells in varios combination as displayed in Table 1B.

Table 1B

Vector + Fragment	Number of colonies	% of colonies with lipase activity
1. Blue 626 + 624	ND	40%
2. Blue 624 + 626	ND	12%
3. Blue 625 + 624	ND	75%
4. Blue 624 + 625	ND	10%

20 Pairwise recombinations of one stop codon mutation on the vector and another on the fragment on the opposite side of the opening site. ND = not determined but a high number.

Row 1 and 2 (in Table 1B) have the mutations located at the same place as row 1 and 2 in Table 1A. As can be seen the number of colonies with lipase activity is clearly higher for the stop codon 25 mutations compared to the frameshift mutations, but the same relative difference between the "mirror image" experiments.

30 This might indicate that the stop codon mutations, which is closer to the "application" of the method, gives a better mixing than frameshift mutations. Row 3 and 4 confirms that the closer the mutation is to the end of the vector the higher chance of mixing.

Recombination with one or two frameshift mutation in the vector

and one or two frameshift mutations in the fragment

Using the same approach as described above the influence of one or two frameshift mutations in the vector and one or two frameshift mutations in the fragment were tested using vectors Blue 425, 426 and 428 (one mutation) and vectors Blue 442, Blue 443 (two mutations) and fragments 442 and 443 (two frameshift mutations) and fragments 424, 425, 426, 427, 428 (one mutation) and wild-type (no mutation)

10

The vectors Blue 442 and 443 are double frameshift mutations: Blue 442 = 428+429 and blue 443 = 427+429 (see Figure 7).

Recombination was performed by transforming 0.5 ml vector (app. 0.1 mg) opened with PstI and 3 ml PCR-fragment (app. 0.5 mg) into 100 ml *Saccharomyces cerevisiae* YNG316 competent cells.

The result of the test is shown in Table 2A and Table 2B

20

Table 2A

Vector + Fragment	Number of colonies	% of colonies with active Lipolase
1. Blue 425 + 442	142	15%
2. Blue 425 + 443	144	14%
3. Blue 426 + 442	42	42%
4. Blue 426 + 443#	77	20%
5. Blue 428 + 443	115	3.8%

One frameshift mutation on the vector and two on the fragment on each side of the opening site. # determined by 6 plates.

25

Table 2B

Vector + Fragment	Number of colonies	% of colonies with active Lipolase
Blue 442 + 424	137	0.5%
Blue 442 + 426	118	1.1%
Blue 442 + 427#	125	1.3%
Blue 443 + 425	540	2.5%

Blue 443 + 426	196	1.5%
Blue 443 + 428	469	3.1%
Blue 442 + wt fragment	135	7.7%
Blue 443 + wt fragment	488	10.7%

Two frameshift mutations on the vector on each side of the opening site and one on the fragment. # determined by 6 plates.

Table 2A shows a rather high number of colonies with lipase activity even with a total of 3 frameshifts (but only one frameshift on the vector) except for the last row where the frameshift on the vector is located far from the opening site. Lane 4 has fewer actives than lane 3 probably due to that the frameshift on the vector is located further away from the opening site than the frameshift on the fragment making the active genes mosaics that are not related to the opening site (see figure 2A). In Table 2B a very low number of actives are observed when there are 2 frameshifts located on the vector. Most of these active colonies are mosaics of the "parent" DNA meaning that the mixing is not related to the opening site (see figure 2B).

#### Recombination with two different vectors or fragments

The result of recombination with two different vectors or fragments the test is shown in Table 3

Table 3

Vector + Fragment	Number of colonies	% of colonies with active Lipolase
Blue 428/pstI + Blue 429/pst #	13	15%
Blue428/pst + Blue 429/PstI + 442	273	4.2%
Blue 442/pstI + 428 + 429	228	0.8%
Blue 443/pstI + 427 + 428	229	1.6%

Recombinations with 2 different vectors or fragments. # Determined by 1 plate.

A low number of colonies are seen for the control experiment in row 1 of table 3 as expected. The fragment added in the middle row has

two frameshifts each corresponding to the frameshift on each vector. Via a tripartite recombination 4.2% actives are created. With two fragments with each one frameshifts and a vector with the same two frameshifts very few actives are found.

5

Recombination with vectors opened at different sites

Opening the vector in one side instead of approximately in the 10 middle still gives good recombination as shown in Table 4. Two vectors opened at different sites can also recombine to some extent (compare with the vector controls in table 13).

Table 4

Vector + Fragment	Number of colonies	% of colonies with active Lipolase
Blue 428/xho + 429	160	11%
Blue 428/xho+Blue 429/pst#	35	6.3%

15 Opening of the vector in one side instead of in the middle. # determined by 6 plates.

20 Recombination at different concentrations of vector and fragment

The relative concentration of vector to fragment do influence the percentage of positive colonies as can be seen in Table 5.

Table 5

Vector + Fragment	Number of colonies	% of colonies with lipase activity
0.5 $\mu$ l Blue 426 + 3 $\mu$ l 442	42	42%
1.5 $\mu$ l Blue 426 + 3 $\mu$ l 442	21	51%
1.5 $\mu$ l Blue 426 + 9 $\mu$ l 442	34	26%
1.5 $\mu$ l Blue 426 + 3 $\mu$ l 427	230	2.8%
1 $\mu$ l Blue 442 + 1 $\mu$ l 425	224	1.16%
1 $\mu$ l Blue 442 + 2 $\mu$ l	429	0.9%

425		
1 $\mu$ l Blue 442 + 4 $\mu$ l 425	434	1.6%
1 $\mu$ l Blue 442 + 8 $\mu$ l 425	481	1.6%
1 $\mu$ l Blue 442 + 16 $\mu$ l 425	497	2.0%

Varying the concentration of the vector or fragment.

5

#### Recombination with fragments of different size

10

Table 6

Vector + Fragment	Number of colonies	% of colonies with active Lipolase
Blue 424 + 425 (260bp)	73	34%
Blue 424 + 425 (489bp)	130	45%
Blue 424 + 424 (480bp)	133	0.3%
Blue 424 + 426 (480bp)	130	36%
Blue 428 + 425 (480bp)	150	28%
Blue 425 + 424 (480bp)	69	0%
Blue 425 + 426 (480bp)	63	55%

Recombination with smaller fragments than 900 bp.

15

#### Recombination with unopened vectors

Transformation with unopened vectors shows a very low degree of recombination (Table 7).

Table 7

Plasmid	Number of colonies	% of colonies with active Lipolase
Blue 428 + Blue 429	887	0.3%

Blue 426 + Blue 425	697	0.7%
---------------------	-----	------

Recombination of unopened plasmids.

Example 4

5

Test of *S. cerevisiae* mutants altered in recombination

Using the same approach as described in Example 3 recombination of opened and unopened vectors and fragments were tested using a 10 *Saccharomyces cerevisiae* rad52 mutant as the recombination host cell. The result is displayed in Table 8.

Table 8

Vector + Fragment	Number of colonies	% of colonies with active Lipolase
Blue 428 + 429	0	0
Blue 442 + 427	0	0
Blue 424 + 425	0	0
Blue 426 + 443	0	0
Plasmid pJSO 37	544	100%

Recombination result in rad52 mutant.

15 The result with rad52 showed that recombination was completely abolished. The RAD52 function is required for classical recombination (but not for unequal sister-strand mitotic recombination) showing that the recombination of opened vector and fragment could involve a classical recombination mechanism.

20

Example 6

25 Recombination of multiple partial overlapping fragments

In order to increase the mixing of the mutations by the recombination method of the invention, recombination of two fragments and one gapped vector were attempted.

30

Table 15

Vector + Fragment	Number of colonies	% of colonies with lipase activity
1. pJSO37/HindIII-XhoI + PCR319+PCR327	> 2000	100%
2. pJSO37/HindIII-XhoI + PCR321+PCR331	≈ 2000	≈ 0.2%

3. pJS037/HindIII-XhoI + PCR319+PCR331	$\approx$ 1500	$\approx$ 1%
4. pJS037/HindIII-XhoI + PCR319+PCR386	> 5000	> 90%
5. pJS037/HindIII-XhoI + PCR321+PCR386	> 5000	$\approx$ 25%
6. Blue 428/HindIII-XhoI + PCR321+PCR331	400	0.2%
7. Blue 428/HindIII-XhoI + PCR319+PCR327	$\approx$ 1500	> 90%
8. Blue 428/HindIII-XhoI + PCR321+PCR327	$\approx$ 150	$\approx$ 10%
9. Blue 428/HindIII-XhoI + PCR327+PCR385	$\approx$ 1500	$\approx$ 10%
10. Blue 429/HindIII-XhoI + PCR319+PCR386	$\approx$ 400	$\approx$ 15%
11. Blue 429/HindIII-XhoI + PCR321+PCR386	$\approx$ 350	$\approx$ 15%
12. Blue 442/HindIII-XhoI + PCR319+PCR327	$\approx$ 1500	$\approx$ 10%
13. Blue 428/HindIII-XhoI +	2	0
14. Blue 429/HindIII-XhoI +	0	0
15. Blue 442/HindIII-XhoI +	6	0
16. Blue 428/HindIII-XhoI + PCR331	4	0
17. Blue 428/HindIII-XhoI + PCR321	2	0

Recombination result of two fragments and a gapped vector. The last 5 rows are controls.

As can be seen in Table 15, the recovery of the *Humicola lanuginosa* 5 lipase gene is very efficient. The last 5 rows in Table 15 shows that the opened vector alone or with only one fragment not covering the whole gap (see figure 3) gives only very few colonies.

The first row is with wild type fragments gives 100% of active 10 colonies.

The second row is with two fragments each containing a frameshift. The fragment PCR331 fragment has the frameshift located at the BglII site which, in this recombination, is not covered by a wild

type fragment (see figure 3) and therefore gives about 0% of active lipase. The same is the case for row 3 and 6.

In the row 4, fragment PCR386 containing a frameshift at the SphI site which is overlapped by wild type sequences in the gapped vector. The frameshift was recombined into less than 10% of the genes which is lower than the result for one fragment recombination in the last row of Table 1A above.

10 In row 5 a rather high mixing is observed between the 2 fragments each containing a frameshift and the wild type gapped vector giving 25% active and 75% inactive lipase colonies. This is probably due to that the fragment PCR321 has the frameshift in the overlap between the 2 fragments and in the gapped region of the vector. If 15 fragment PCR386 contributes to 10% inactives like in row 4, fragment PCR321 gives the remaining 65% inactives - therefore PCR386 gives 35% wt in the overlap.

20 Row 7 is the "mirror image" of row 4 with the frameshift at the SphI site on the vector (see Figure 7) and 2 wild type fragments giving an integration of the wild type fragment into more than 90% of the vectors.

25 Row 8 shows like in row 5 that the frameshift of PCR321 in the overlap and gap region gives a very high number of inactive.

In row 9, fragment PCR385 with a frameshift in the vector overlap, causes a very high number of inactives.

30 Row 10 gives a rather high number of inactives compared to row 7 and 4. It is not increased in row 11.

Row 12 shows that two frameshifts on the vector gives a lower 35 number of actives compared to one in row 7.

The recombination of 3 partial overlapping fragments into a gapped vector is also very efficient as seen in Table 16. The last row with the vector alone gives very few colonies. As can be seen in figure 4 all fragments used are wt. In the first row in table 16, 40 there are rather long overlaps between the vector and fragments, but in the middle row the overlap between PCR353 and 355 is only 10

bp long and it is still very efficiently recombined! This surprising result may be utilized for very easy domain shuffling of even distantly related genes. For example can 3 different domains from 10 different genes be made as PCR fragments, designed to have 5 a 10 to 20 bp overlap by primer design and recombined together and subsequently screened for the best combination (1000 possible combinations).

Table 16

Vector + Fragment	Number of colonies	% of colonies with active Lipolase
pJS037/PvuII-SpeI + PCR353+PCR354+PCR367	> 5000	100%
pJS037/PvuII-SpeI + PCR353+PCR355+PCR367	> 5000	100%
pJS037/PvuII-SpeI	20	100%

10 Recombination result of 3 fragments and a gapped vector. The last row is a control.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

5 (i) APPLICANT:  
(A) NAME: Novo Nordisk A/S  
(B) STREET: Novo Alle  
(C) CITY: Bagsvaerd  
10 (E) COUNTRY: Denmark  
(F) POSTAL CODE (ZIP): DK-2880  
(G) TELEPHONE: +45 4444 8888  
(H) TELEFAX: +45 4449 3256

(ii) TITLE OF INVENTION: Method for preparing polypeptide variants

15 (iii) NUMBER OF SEQUENCES: 15

(iv) COMPUTER READABLE FORM:  
(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
20 (D) SOFTWARE: PatentIn Release #1.0, Version #1.308 (EPO)

## (2) INFORMATION FOR SEQ ID NO: 1:

25 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "Primer 2843"  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ACATACATTA CGTGACGGG

20

35 (2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

40 (ii) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "Primer 4699"

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CGGTACCCGG GGATCCAC

18

50 (2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

55 (ii) MOLECULE TYPE: other nucleic acid  
(A) DESCRIPTION: /desc = "Primer 5164"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

60 AATTACATCA TGCAGCCC

18

## (2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
5 (ii) MOLECULE TYPE: other nucleic acid  
      (A) DESCRIPTION: /desc = "Primer 8487"  
      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CATTIGCTCC GGCTGCAGGG A

21

10 (2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:  
      (A) LENGTH: 60 base pairs  
15 (B) TYPE: nucleic acid  
      (C) STRANDEDNESS: single  
      (D) TOPOLOGY: linear  
      (ii) MOLECULE TYPE: other nucleic acid  
      (A) DESCRIPTION: /desc = "Primer 4548"  
20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GTGTTCCGCC GGTCTGTACG GTCAGGAATT CTGCAAAAGC  
CCTGTTTCCG A<sup>T</sup>CTGGGGGG

60

25 (2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:  
      (A) LENGTH: 21 base pairs  
30 (B) TYPE: nucleic acid  
      (C) STRANDEDNESS: single  
      (D) TOPOLOGY: linear  
      (ii) MOLECULE TYPE: other nucleic acid  
      (A) DESCRIPTION: /desc = "Primer 5576"  
      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GGTCTGTACG GTCAGGAATT C

21

(2) INFORMATION FOR SEQ ID NO: 7:

40 (i) SEQUENCE CHARACTERISTICS:  
      (A) LENGTH: 19 base pairs  
      (B) TYPE: nucleic acid  
      (C) STRANDEDNESS: single  
      (D) TOPOLOGY: linear  
45 (ii) MOLECULE TYPE: other nucleic acid  
      (A) DESCRIPTION: /desc = "Primer 5578"  
      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

CGTTTCGGGT GACGGGGAC

19

50 (2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:  
      (A) LENGTH: 18 base pairs  
55 (B) TYPE: nucleic acid  
      (C) STRANDEDNESS: single  
      (D) TOPOLOGY: linear  
      (ii) MOLECULE TYPE: other nucleic acid  
      (A) DESCRIPTION: /desc = "Primer 1596"  
60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGAGCAATTG TCATTTAT

18

(2) INFORMATION FOR SEQ ID NO: 9:

5

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 64 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
  - (A) DESCRIPTION: /desc = "Primer 4545"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GCATTGGCAA CTGTTGCCGG AGCAGACCTG CGTGGAAATG  
GGTATGATAT CGACGGTGTTC TCAT

64

15 (2) INFORMATION FOR SEQ ID NO: 10:

15 (i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 876 base pair

15 (B) TYPE: nucleic acid

15 (C) STRANDEDNESS: single

15 (D) TOPOLOGY: circular

20 (ii) MOLECULE TYPE: other nucleic acid

20 (A) DESCRIPTION: /desc

25 (vi) ORIGINAL SOURCE:

25 (B) STRAIN: Humicola lanuginosus

30 (ix) FEATURE:

30 (A) NAME/KEY: CDS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

35 ATG AGG AGC TCC CTT GTG CTG TTC TTT GTC TCT GCG TGG ACG GCC TTG  
Met Arg Ser Ser Leu Val Leu Phe Phe Val Ser Ala Trp Thr Ala Leu

68

```

40  GCC AGT CCT ATT CGT CGA GAG GTC TCG CAG GAT CTG TTT AAC CAG TTC
     Ala Ser Pro Ile Arg Arg Glu Val Ser Gln Asp Leu Phe Asn Gln Phe
          20          25          30

```

96

45 AAT CTC TTT GCA CAG TAT TCT GCA GCC GCA TAC TGC GGA AAA AAC AAT  
 Asn Leu Phe Ala Gln Tyr Ser Ala Ala Ala Tyr Cys Gly Lys Asn Asn  
 35 40 45

144

GAT GCC CCA GCT GGT ACA AAC ATT ACG TGC ACG GGA AAT GCC TGC CCC  
 Asp Ala Pro Ala Gly Thr Asn Ile Thr Cys Thr Gly Asn Ala Cys Pro  
 50 55 60

192

GAG	GTA	GAG	AAG	GCG	GAT	GCA	ACG	TTT	CTC	TAC	TCG	TTT	GAA	GAC	TCT
Glu	Val	Glu	Lys	Ala	Asp	Ala	Thr	Phe	Leu	Tyr	Ser	Phe	Glu	Asp	Ser
65		70							75					80	

240

55 GGA GIG GGC GAT GTC ACC GGC TTC CTT GCT CTC GAC AAC ACG AAC AAA  
 Gly Val Gly Asp Val Thr Gly Phe Leu Ala Leu Asp Asn Thr Asn Lys  
 85 90 95

283

60 Leu Ile Val Leu Ser Phe Arg Gly Ser Arg Ser Ile Glu Asn Trp Ile  
 100 105 110  
 GGG AAT CTT AAC TTC GAC TTG AAA GAA ATA AAT GAC ATT TGC TCC GGC  
 Gly Asn Leu Asn Phe Asp Leu Lys Glu Ile Asn Asp Ile Cys Ser Gly  
 65 115 120 125

	TGC AGG GGA CAT GAC GGC TTC ACT TCG TCC TGG AGG TCT GTA GCC GAT	432
	Cys Arg Gly His Asp Gly Phe Thr Ser Ser Trp Arg Ser Val Ala Asp	
	130 135 140	
5	ACG TTA AGG CAG AAG GTG GAG GAT GCT GTG AGG GAG CAT CCC GAC TAT	480
	Thr Leu Arg Gln Lys Val Glu Asp Ala Val Arg Glu His Pro Asp Tyr	
	145 150 155 160	
10	CGC GTG GTG TTT ACC GGA CAT AGC TTG GGT GGT GCA TTG GCA ACT GTT	528
	Arg Val Val Phe Thr Gly His Ser Leu Gly Gly Ala Leu Ala Thr Val	
	165 170 175	
15	GCC GGA GCA GAC CTG CGT GGA AAT GGG TAT GAT ATC GAC GTG TTT TCA	576
	Ala Gly Ala Asp Leu Arg Gly Asn Gly Tyr Asp Ile Asp Val Phe Ser	
	180 185 190	
20	TAT GGC GCC CCC CGA GTC GGA AAC AGG GCT TTT GCA GAA TTC CTG ACC	624
	Tyr Gly Ala Pro Arg Val Gly Asn Arg Ala Phe Ala Glu Phe Leu Thr	
	195 200 205	
25	GTA CAG ACC GGC GGA ACA CTC TAC CGC ATT ACC CAC ACC AAT GAT ATT	672
	Val Gln Thr Gly Gly Thr Leu Tyr Arg Ile Thr His Thr Asn Asp Ile	
	210 215 220	
30	GTC CCT AGA CTC CCG CCG CGC GAA TTC GGT TAC AGC CAT TCT AGC CCA	720
	Val Pro Arg Leu Pro Pro Arg Glu Phe Gly Tyr Ser His Ser Ser Pro	
	225 230 235 240	
35	GAG TAC TGG ATC AAA TCT GGA ACC CTT GTC CCC GTC ACC CGA AAC GAT	768
	Glu Tyr Trp Ile Lys Ser Gly Thr Leu Val Pro Val Thr Arg Asn Asp	
	245 250 255	
40	ATC GTG AAG ATA GAA GGC ATC GAT GCC ACC GGC GGC AAT AAC CAG CCT	816
	Ile Val Lys Ile Glu Gly Ile Asp Ala Thr Gly Gly Asn Asn Gln Pro	
	260 265 270	
45	AAC ATT CCG GAT ATC CCT GCG CAC CTA TGG TAC TTC GGG TTA ATT GGG	864
	Asn Ile Pro Asp Ile Pro Ala His Leu Trp Tyr Phe Gly Leu Ile Gly	
	275 280 285	
50	ACA TGT CTT TAG	376
	Thr Cys Leu	
	290	
55	(2) INFORMATION FOR SEQ ID NO: 11:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 292 amino acids	
	(B) TYPE: amino acid	
	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: protein	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:	
60	Met Arg Ser Ser Leu Val Leu Phe Phe Val Ser Ala Trp Thr Ala Leu	
	1 5 10 15	
	Ala Ser Pro Ile Arg Arg Glu Val Ser Gln Asp Leu Phe Asn Gln Phe	
	20 25 30	
	Asn Leu Phe Ala Gln Tyr Ser Ala Ala Ala Tyr Cys Gly Lys Asn Asn	
	35 40 45	
65	Asp Ala Pro Ala Gly Thr Asn Ile Thr Cys Thr Gly Asn Ala Cys Pro	
	50 55 60	

Glu Val Glu Lys Ala Asp Ala Thr Phe Leu Tyr Ser Phe Glu Asp Ser  
 65 70 75 80  
 Gly Val Gly Asp Val Thr Gly Phe Leu Ala Leu Asp Asn Thr Asn Lys  
 5 85 90 95  
 Leu Ile Val Leu Ser Phe Arg Gly Ser Arg Ser Ile Glu Asn Trp Ile  
 100 105 110  
 Gly Asn Leu Asn Phe Asp Leu Lys Glu Ile Asn Asp Ile Cys Ser Gly  
 10 115 120 125  
 Cys Arg Gly His Asp Gly Phe Thr Ser Ser Trp Arg Ser Val Ala Asp  
 130 135 140  
 15 Thr Leu Arg Gln Lys Val Glu Asp Ala Val Arg Glu His Pro Asp Tyr  
 145 150 155 160  
 Arg Val Val Phe Thr Gly His Ser Leu Gly Gly Ala Leu Ala Thr Val  
 20 165 170 175  
 Ala Gly Ala Asp Leu Arg Gly Asn Gly Tyr Asp Ile Asp Val Phe Ser  
 180 185 190  
 25 Tyr Gly Ala Pro Arg Val Gly Asn Arg Ala Phe Ala Glu Phe Leu Thr  
 195 200 205  
 Val Gin Thr Gly Gly Thr Leu Tyr Arg Ile Thr His Thr Asn Asp Ile  
 210 215 220  
 30 Val Pro Arg Leu Pro Pro Arg Glu Phe Gly Tyr Ser His Ser Ser Pro  
 225 230 235 240  
 Glu Tyr Trp Ile Lys Ser Gly Thr Leu Val Pro Val Thr Arg Asn Asp  
 35 245 250 255  
 Ile Val Lys Ile Glu Gly Ile Asp Ala Thr Gly Gly Asn Asn Gln Pro  
 260 265 270  
 40 Asn Ile Pro Asp Ile Pro Ala His Leu Trp Tyr Phe Gly Leu Ile Gly  
 275 280 285  
 Thr Cys Leu  
 290  
 45

## (2) INFORMATION FOR SEQ ID NO: 12:

50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 876 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: circular

55 (ii) MOLECULE TYPE: other nucleic acid  
 (A) DESCRIPTION: /desc = "Vector pJSG37"

60 (vi) ORIGINAL SOURCE:  
 (B) STRAIN: *Humicola lanuginosa*

65 (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 1..876

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

ATG	AGG	AGC	TCC	CTT	GTG	CTG	TTC	TTT	GTC	TCT	GCG	TGG	ACG	GCC	TTG	48	
Met	Arg	Ser	Ser	Leu	Val	Leu	Phe	Phe	Val	Ser	Ala	Trp	Thr	Ala	Leu		
1				5					10					15			
5	GCC	AGT	CCT	ATA	CGT	AGA	GAG	GTC	TCG	CAG	GAT	CTG	TTT	AAC	CAG	TTC	96
	Ala	Ser	Pro	Ile	Arg	Arg	Glu	Val	Ser	Gln	Asp	Leu	Phe	Asn	Gln	Phe	
				20					25					30			
10	AAT	CTC	TTT	GCA	CAG	TAT	TCA	GCT	GCC	GCA	TAC	TGC	GGA	AAA	AAC	AAT	144
	Asn	Leu	Phe	Ala	Gln	Tyr	Ser	Ala	Ala	Ala	Tyr	Cys	Gly	Lys	Asn	Asn	
				35					40					45			
15	GAT	GCC	CCA	GCA	GGT	ACA	AAC	ATT	ACG	TGC	ACG	GGA	AAT	GCA	TGC	CCC	192
	Asp	Ala	Pro	Ala	Gly	Thr	Asn	Ile	Thr	Cys	Thr	Gly	Asn	Ala	Cys	Pro	
				50					55					60			
20	GAG	GTA	GAG	AAG	GCG	GAT	GCA	ACG	TTT	CTC	TAC	TCG	TTT	GAA	GAC	TCT	240
	Glu	Val	Glu	Lys	Ala	Asp	Ala	Thr	Phe	Leu	Tyr	Ser	Phe	Glu	Asp	Ser	
				65					70					75			
														80			
25	GGA	GTG	GGC	GAT	GTC	ACC	GGC	TTC	CTT	GCT	CTC	GAC	AAC	ACG	AAC	AAG	288
	Gly	Val	Gly	Asp	Val	Thr	Gly	Phe	Leu	Ala	Leu	Asp	Asn	Thr	Asn	Lys	
									85					90			
														95			
30	CTT	ATC	GTC	CTC	TCT	TTC	CGT	GGC	TCA	AGA	TCT	ATA	GAG	AAC	TGG	ATC	336
	Leu	Ile	Val	Leu	Ser	Phe	Arg	Gly	Ser	Arg	Ser	Ile	Glu	Asn	Trp	Ile	
									100					105			
														110			
35	GGG	AAT	CTT	AAC	TTC	GAC	TTG	AAA	GAA	ATA	AAT	GAC	ATT	TGC	TCC	GGC	384
	Gly	Asn	Leu	Asn	Phe	Asp	Leu	Lys	Glu	Ile	Asn	Asp	Ile	Cys	Ser	Gly	
									115					120			
														125			
40	TGC	AGG	GGA	CAT	GAC	GGC	TTC	ACT	TGG	TCC	TGG	AGG	TCT	GTA	GCC	GAT	432
	Cys	Arg	Gly	His	Asp	Gly	Phe	Thr	Ser	Ser	Trp	Arg	Ser	Val	Ala	Asp	
									130					135			
														140			
45	ACG	TTA	AGG	CAG	AAG	GTG	GAG	GAT	GCT	GTT	CGC	GAG	CAT	CCC	GAC	TAT	480
	Thr	Leu	Arg	Gly	Val	Gly	Asp	Ala	Val	Arg	Glu	His	Pro	Asp	Tyr		
									145					150			
														155			
														160			
50	CGC	GTG	GTG	TTT	ACC	GGC	CAT	AGC	CTT	GGT	GGT	GCG	CTA	GCA	ACT	GTT	528
	Arg	Val	Val	Phe	Thr	Gly	His	Ser	Ile	Gly	Gly	Ala	Leu	Ala	Thr	Val	
									165					170			
														175			
55	GCC	GGA	GCA	GAC	CTG	CGT	GGG	AAT	GGG	TAT	GAT	ATC	GAC	GTG	TTT	TCA	576
	Ala	Gly	Ala	Asp	Leu	Arg	Gly	Asn	Gly	Tyr	Asp	Ile	Asp	Val	Phe	Ser	
									180					185			
														190			
60	TAT	GGC	GCC	CCC	CGA	GTC	GGT	AAC	CGT	GCT	TTT	GCA	GAA	TTC	CTG	ACC	624
	Tyr	Gly	Ala	Pro	Arg	Val	Gly	Asn	Arg	Ala	Phe	Ala	Glu	Phe	Leu	Thr	
									195					200			
														205			
65	GTA	CAG	ACC	GGC	GGT	ACC	CTC	TAC	CGC	ATT	ACC	CAC	ACC	AAT	GAT	ATT	672
	Val	Gln	Thr	Gly	Gly	Thr	Leu	Tyr	Arg	Ile	Thr	His	Thr	Asn	Asp	Ile	
									210					215			
														220			
70	GTC	CCT	AGA	CTC	CCG	CCT	CGA	GAA	TTC	GGT	TAC	AGC	CAT	TCT	AGC	CCA	720
	Val	Pro	Arg	Leu	Pro	Pro	Arg	Glu	Phe	Gly	Tyr	Ser	His	Ser	Ser	Pro	
									225					230			
														235			
75	GAG	TAC	TGG	ATC	AAA	TCT	GGA	ACA	CTA	GTC	CCC	GTC	ACC	CGA	AAC	GAT	768
	Glu	Tyr	Trp	Ile	Lys	Ser	Gly	Thr	Leu	Val	Pro	Val	Thr	Arg	Asn	Asp	
									245					250			
														255			
80	ATC	GTG	AAG	ATA	GAA	GGC	ATC	GAT	GCC	ACC	GGC	GGC	AAT	AAC	CAG	CCT	816
	Ile	Val	Lys	Ile	Glu	Gly	Ile	Asp	Ala	Thr	Gly	Gly	Asn	Asn	Gln	Pro	
									260					265			
														270			

AAC ATT CCG GAT ATC CCT GCG CAC CTA TGG TAC TTC GGG TTA ATT GGG  
 Asn Ile Pro Asp Ile Pro Ala His Leu Trp Tyr Phe Gly Leu Ile Gly  
 275 280 285 864

5

ACA TGT CTT TAG  
 Thr Cys Leu \* 876  
 290

10

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 292 amino acids  
 15 (B) TYPE: amino acid  
 (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

20 Met Arg Ser Ser Leu Val Leu Phe Phe Val Ser Ala Trp Thr Ala Leu  
 1 5 10 15

25 Ala Ser Pro Ile Arg Arg Glu Val Ser Gln Asp Leu Phe Asn Gln Phe  
 20 25 30

Asn Leu Phe Ala Gln Tyr Ser Ala Ala Ala Tyr Cys Gly Lys Asn Asn  
 35 40 45

30 Asp Ala Pro Ala Gly Thr Asn Ile Thr Cys Thr Gly Asn Ala Cys Pro  
 50 55 60

Glu Val Glu Lys Ala Asp Ala Thr Phe Leu Tyr Ser Phe Glu Asp Ser  
 65 70 75 80

35 Gly Val Gly Asp Val Thr Gly Phe Leu Ala Leu Asp Asn Thr Asn Lys  
 85 90 95

40 Leu Ile Val Leu Ser Phe Arg Gly Ser Arg Ser Ile Glu Asn Trp Ile  
 100 105 110

Gly Asn Leu Asn Phe Asp Leu Lys Glu Ile Asn Asp Ile Cys Ser Gly  
 115 120 125

45 Cys Arg Gly His Asp Gly Phe Thr Ser Ser Trp Arg Ser Val Ala Asp  
 130 135 140

50 Thr Leu Arg Gln Lys Val Glu Asp Ala Val Arg Glu His Pro Asp Tyr  
 145 150 155 160

Arg Val Val Phe Thr Gly His Ser Leu Gly Gly Ala Leu Ala Thr Val  
 165 170 175

55 Ala Gly Ala Asp Leu Arg Gly Asn Gly Tyr Asp Ile Asp Val Phe Ser  
 180 185 190

Tyr Gly Ala Pro Arg Val Gly Asn Arg Ala Phe Ala Glu Phe Leu Thr  
 195 200 205

60 Val Gln Thr Gly Gly Thr Leu Tyr Arg Ile Thr His Thr Asn Asp Ile  
 210 215 220

Val Pro Arg Leu Pro Pro Arg Glu Phe Gly Tyr Ser His Ser Ser Pro  
 225 230 235 240

65 Glu Tyr Trp Ile Lys Ser Gly Thr Leu Val Pro Val Thr Arg Asn Asp  
 245 250 255

Ile Val Lys Ile Glu Gly Ile Asp Ala Thr Gly Gly Asn Asn Gln Pro  
 260 265 270  
 5 Asn Ile Pro Asp Ile Pro Ala His Leu Trp Tyr Phe Gly Leu Ile Gly  
 275 280 285  
 Thr Cys Leu  
 290  
 10

## (2) INFORMATION FOR SEQ ID NO: 14:

15 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 864 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:  
 (B) STRAIN: *Pseudomonas* sp.

25 (ix) FEATURE:  
 (A) NAME/KEY: *mat\_peptide*  
 (B) LOCATION: 1..864

30 (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 1..864

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

35	TTC GGC TCC TCG AAC TAC ACC AAG ACC CAG TAC CCG ATC GTC CTG ACC Phe Gly Ser Ser Asn Tyr Thr Lys Thr Gln Tyr Pro Ile Val Leu Thr 1 5 10 15	48
40	CAC GGC ATG CTC GGT TTC GAC AGC CTG CTT GGA GTC GAC TAC TGG TAC His Gly Met Leu Gly Phe Asp Ser Leu Leu Gly Val Asp Tyr Trp Tyr 20 25 30	96
45	GGC ATT CCC TCA GCC CTG CGT AAA GAC GGC GCC ACC GTC TAC GTC ACC Gly Ile Pro Ser Ala Leu Arg Lys Asp Gly Ala Thr Val Tyr Val Thr 35 40 45	144
50	GAA GTC AGC CAG CTC GAC ACC TCC GAA GCC CGA GGT GAG CAA CTG CTG Glu Val Ser Gln Leu Asp Thr Ser Glu Ala Arg Gly Glu Gln Leu Leu 50 55 60	192
55	ACC CAA GTC GAG GAA ATC GTG GCC ATC AGC GGC AAG CCC AAG GTC AAC Thr Gln Val Glu Glu Ile Val Ala Ile Ser Gly Lys Pro Lys Val Asn 65 70 75 80	240
60	CTG TTC GGC CAC AGC CAT GGC GGG CCT ACC ATC CGC TAC GTT GCC GCC Leu Phe Gly His Ser His Gly Gly Pro Thr Ile Arg Tyr Val Ala Ala 85 90 95	288
65	GTG CGC CCG GAT CTG GTC GCC TCG GTC ACC AGC ATT GGC GCG CCG CAC Val Arg Pro Asp Leu Val Ala Ser Val Thr Ser Ile Gly Ala Pro His 100 105 110	336
70	AAG GGT TCG GCC ACC GCC GAC TTC ATC CGC CAG GTG CCG GAA GGA TCG Lys Gly Ser Ala Thr Ala Asp Phe Ile Arg Gln Val Pro Glu Gly Ser 115 120 125	384

	GCC AGC GAA GCG ATT CTG GCC GGG ATC GTC AAT GGT CTG GGT GCG CTG Ala Ser Glu Ala Ile Leu Ala Gly Ile Val Asn Gly Leu Gly Ala Leu 130 135 140	432
5	ATC AAC TTC CTT TCC GGC AGC AGT TCG GAC ACC CCA CAG AAC TCG CTG Ile Asn Phe Leu Ser Gly Ser Ser Asp Thr Pro Gln Asn Ser Leu 145 150 155 160	480
10	GGC ACG CTG GAG TCA CTG AAC TCC GAA GGC GCC GCA CGG TTT AAC GCC Gly Thr Leu Glu Ser Leu Asn Ser Glu Gly Ala Ala Arg Phe Asn Ala 165 170 175	528
15	CGC TTC CCC CAG GGG GTA CCA ACC AGC GCC TGC GGC GAG GGC GAT TAC Arg Phe Pro Gln Gly Val Pro Thr Ser Ala Cys Gly Glu Gly Asp Tyr 180 185 190	576
20	GTG GTC AAT GGC GTG CGC TAT TAC TCC TGG AGG GGC ACC AGC CCG CTG Val Val Asn Gly Val Arg Tyr Tyr Ser Trp Arg Gly Thr Ser Pro Leu 195 200 205	624
25	ACC AAC GTA CTC GAC CCC TCC GAC CTG CTG CTC GGC GCC ACC TCC CTG Thr Asn Val Leu Asp Pro Ser Asp Leu Leu Leu Gly Ala Thr Ser Leu 210 215 220	672
30	CGG CTG GGT ATG GTG ATC CGC GAC AAC TAC CGG ATG AAC CAC CTG GAC Arg Leu Gly Met Val Ile Arg Asp Asn Tyr Arg Met Asn His Leu Asp 245 250 255	768
35	GAG GTG AAC CAG ACC TTC GGG CTG ACC AGC ATC TTC GAG ACC AGC CCG Glu Val Asn Gln Thr Phe Gly Leu Thr Ser Ile Phe Glu Thr Ser Pro 260 265 270	816
40	GTA TCG GTC TAT CGC CAG CAA GCC AAT CGC CTG AAG AAC GCC GGG CTC Val Ser Val Tyr Arg Gln Gln Ala Asn Arg Leu Lys Asn Ala Gly Leu 275 280 285	864
(2) INFORMATION FOR SEQ ID NO: 15:		
45	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 288 amino acids (B) TYPE: amino acid (D) TOPOLOGY: linear	
50	(ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15: Phe Gly Ser Ser Asn Tyr Thr Lys Thr Gln Tyr Pro Ile Val Leu Thr 1 5 10 15	
55	His Gly Met Leu Gly Phe Asp Ser Leu Leu Gly Val Asp Tyr Trp Tyr 20 25 30	
60	Gly Ile Pro Ser Ala Leu Arg Lys Asp Gly Ala Thr Val Tyr Val Thr 35 40 45	
65	Glu Val Ser Gln Leu Asp Thr Ser Glu Ala Arg Gly Glu Gln Leu Leu 50 55 60	
70	Thr Gln Val Glu Glu Ile Val Ala Ile Ser Gly Lys Pro Lys Val Asn 65 70 75 80	
75	Leu Phe Gly His Ser His Gly Gly Pro Thr Ile Arg Tyr Val Ala Ala 85 90 95	

Val Arg Pro Asp Leu Val Ala Ser Val Thr Ser Ile Gly Ala Pro His  
100 105 110

5 Lys Gly Ser Ala Thr Ala Asp Phe Ile Arg Gln Val Pro Glu Gly Ser  
115 120 125

Ala Ser Glu Ala Ile Leu Ala Gly Ile Val Asn Gly Leu Gly Ala Leu  
130 135 140

10 Ile Asn Phe Leu Ser Gly Ser Ser Asp Thr Pro Gln Asn Ser Leu  
145 150 155 160

Gly Thr Leu Glu Ser Leu Asn Ser Glu Gly Ala Ala Arg Phe Asn Ala  
15 165 170 175

Arg Phe Pro Gln Gly Val Pro Thr Ser Ala Cys Gly Glu Gly Asp Tyr  
180 185 190

20 Val Val Asn Gly Val Arg Tyr Tyr Ser Trp Arg Gly Thr Ser Pro Leu  
195 200 205

Thr Asn Val Leu Asp Pro Ser Asp Leu Leu Leu Gly Ala Thr Ser Leu  
210 215 220

25 Thr Phe Gly Phe Glu Ala Asn Asp Gly Leu Val Gly Arg Cys Ser Ser  
225 230 235 240

Arg Leu Gly Met Val Ile Arg Asp Asn Tyr Arg Met Asn His Leu Asp  
30 245 250 255

Glu Val Asn Gln Thr Phe Gly Leu Thr Ser Ile Phe Glu Thr Ser Pro  
260 265 270

35 Val Ser Val Tyr Arg Gln Gln Ala Asn Arg Leu Lys Asn Ala Gly Leu  
275 280 285

## PATENT CLAIMS

1. A method for preparing polypeptide variants by shuffling different nucleotide sequences of homologous DNA sequences by *in vivo* recombination comprising the steps of
  - a) forming at least one circular plasmid comprising a DNA sequence encoding a polypeptide,
  - b) opening said circular plasmid(s) within the DNA sequence(s) encoding the polypeptide(s),
  - c) preparing at least one DNA fragment comprising a DNA sequence homologous to at least a part of the polypeptide coding region on at least one of the circular plasmid(s),
  - d) introducing at least one of said opened plasmid(s), together with at least one of said homologous DNA fragment(s) covering full-length DNA sequences encoding said polypeptide(s) or parts thereof, into a recombination host cell,
  - e) cultivating said recombination host cell, and
  - f) screening for positive polypeptide variants.
2. The method according to claim 1, wherein more than one cycle of step a) to f) are performed.
3. The method according to claims 1 and 2, wherein two or more opened plasmids are shuffled with one or more homologous DNA fragments in the same shuffling cycle.
4. The method according to any of claims 1 to 3, wherein the opened plasmid(s) is(are) gapped.
5. The method according to any of claims 1 to 4 wherein the ratio between the opened plasmid(s) and homologous DNA fragment(s) are in the range from 20:1 to 1:50, preferably from 2:1 to 1:10 (mol vector:mol fragments) with the specific concentrations being from 1 pM to 10 M of the DNA.
6. The method according to any claims 1 to 5, wherein 2 or more, preferably from 2 to 6, especially 2 to 4 of the DNA fragments have partially overlapping regions.

7. The method according to claim 6, wherein the overlapping regions of the DNA fragments lies in the range from 5 to 5000 bp, preferably from 10 bp to 500 bp, especially 10 bp to 100 bp.

5 8. The method according to any of claims 1 and 8, wherein at least one cycle of step a) to f) is backcrossing with the initially used DNA fragments.

9. The method according to any of claims 1 and 8, wherein the 10 plasmid(s) is(are) opened in the region around the middle of the DNA sequence(s) encoding the polypeptide(s).

10. The method according to any of claims 1 to 9, wherein the 15 plasmid(s) is(are) opened close to a mutation in the DNA sequence(s) encoding the polypeptide(s).

11. The method according to any of claims 1 to 10, wherein the DNA fragment(s) prepared in step c) is(are) prepared under conditions suitable for high, medium or low mutagenesis.

20 12. The method according to any of claims 1 to 11, wherein the polypeptides producible from the input DNA sequences are enzymes or proteins with biological activity.

25 13. The method according to claim 12, wherein the polypeptides are enzymes selected from the group including proteases, lipases, cutinases, cellulases, amylases, peroxidases, oxidases and phytases.

30 14. The method according to claim 12, wherein the polypeptides are proteins with biological activity selected from the group including insulin, ACTH, glucagon, somatostatin, somatotropin, thymosin, parathyroid hormone, pigmentary hormones, somatomedin, erythropoietin, luteinizing hormone, chorionic gonadotropin, hypothalamic releasing factors, antidiuretic hormones, thyroid stimulating 35 hormone, relaxin, interferon, thrombopoietin (TPO) and prolactin.

15. The method according to any of claims 1 to 13, wherein at least one of the initially used input DNA sequences is a wild-type DNA sequence, such as a DNA sequence coding for wild-type enzymes, in 40 particular lipases, derived from filamentous fungi, such as *Humicola* sp., in particular *Humicola lanuginosa*, especially *Humicola*

*lanuginosa* DSM 4109.

16. The method according to claim 15, wherein at least one of the input DNA sequences is selected from the group of vectors (a) to (f) 5 and/or DNA fragments (g) to (aa) coding for *Humicola lanuginosa* lipase variants.

17. The method according to any of claims 1 to 13, wherein at least one of the initially used input DNA sequences is a wild-type DNA 10 sequence, such as a DNA sequence coding for wild-type enzymes, in particular lipases, derived from filamentous fungi of the genera *Absidia*, *Rhizopus*, *Emericella*, *Aspergillus*, *Penicillium*, *Eupenicillium*, *Paecilomyces*, *Talaromyces*, *Thermoascus* and *Sclerocleista*.

15 18. The method according to any of claims 1 and 13, wherein at least one of the initially used input DNA sequences is a wild-type DNA sequence, such as a DNA sequence coding for wild-type enzymes, in particular lipases, derived from bacteria, such as *Pseudomonas* sp., 20 in particular *Ps. fragi*, *Ps. stutzeri*, *Ps. cepacia*, *Ps. fluorescens*, *Ps. plantarii*, *Ps. gladioli*, *Ps. alcaligenes*, *Ps. pseudoalcaligenes*, *Ps. mendocina*, *Ps. auroginosa*, *Ps. glumae*, *Ps. syringae*, *Ps. wisconsinensis*, or a strain of *Bacillus* sp., in particular *B. subtilis*, *B. stearothermophilus* or *B. pumilus*, or 25 or a strain of *Streptomyces* sp., in particular *S. scabies*, or a strain of *Chromobacterium* sp. in particular *C. viscosum*.

19. The method according to any of claims 1 to 13, wherein at least 30 one of the initially used input DNA sequences is a variant DNA sequence, such as a DNA sequence coding for a variant enzyme, in particular lipase variants, derived from yeasts, such as *Candida* sp., in particular *Candida rugosa* or *Geotrichum* sp., in particular *Geotrichum candidum*.

35 20. The method according to any of claims 1 to 19, wherein the homologous input DNA sequences are at least 60%, preferably at least 70%, better more than 80%, especially more than 90%, and even up to 100% homologous.

40 21. The method according to any of claims 1 to 20, wherein the

recombination host cell is a eukaryotic cell, such as a fungal cell or a plant cell.

22. The method according to claim 21, wherein said fungal cell is a yeast cell from the group of cell of *Saccharomyces* sp., in particular strains of *Saccharomyces cerevisiae* or *Saccharomyces kluyveri* or *Schizosaccharomyces* sp., in particular *Schizosaccharomyces pombe*, or *Kluyveromyces* sp., such as *K. lactis*, or *Hansenula* sp., in particular *H. polymorpha*, or *Pichia* sp., in particular *P. pastoris*, or a filamentous fungi from the group of *Aspergillus* sp., in particular *A. niger*, *A. nidulans* or *A. oryzae*, or *Neurospora* sp., or *Fusarium* sp., in particular *F. oxysporum*, or *Trichoderma* sp..

15 23. The method according to any of claims 1 to 22, wherein the plasmid DNA sequence(s) coding for the polypeptide(s) is(are) operably linked to a replication sequence.

20 24. The method according to claim 23, wherein the plasmid DNA sequence(s) encoding the polypeptide(s) is(are) operably linked to a functional promoter sequence.

25 25. The method according to claim 24, wherein the plasmid is an expression plasmid.

26. The method according to claim 25, wherein the expression plasmid is pJS026 or pJS037.

Title: Method for preparing polypeptide variants

1/10

5447 ATGAGGAGCTCCCTGTGCTCTTGTCTCGTGGACGGCCTGGCCAGTCCTATT 5506  
 1 M R S S L V L F F V S A W T A L A S P I -  
 5507 CGTCGAGAGGTCTCGCAGGATCTGTTAACCGATCTCAATCTCTTGACAGTATTCTGCA 5566  
 21 R R E V S Q D L F N Q F N L F A Q Y S A -  
 5567 GCCGCATACTGCGGAAAAAACAAATGATGCCAGCTGGTACAAACATTACGTGACGGGA 5626  
 41 A A Y C G K N N D A P A G T N I T C T G -  
 5627 AATGCCTGCCCGAGGTAGAGAAGGCGGATGCAACGTTCTACTCGTTGAAGACTCT 5686  
 61 N A C P E V E K A D A T F L Y S F E D S -  
 5687 GGAGTGGCGATGTACCGGCTTCCTGCTCTCGACAAACACGAACAAATTGATCGTCTC 5746  
 81 G V G D V T G F L A L D N T N K L I V L -  
 5747 TCTTCCGTGGCTCTCGTCCATAGAGAACTGGATGGAACTTAACTTCGACTTGAAA 5806  
 101 S F R G S R S I E N W I G N L N F D L K -  
 5807 GAAATAAAATGACATTGCTCCGGCTGCAGGGGACATGACGGCTTCACTCGTCCTGGAGG 5866  
 121 E I N D I C S G C R G H D G F T S S W R -  
 5867 TCTGTAGCCGATACTTAAGGCAGAAGGTGGAGGATGCTGTGAGGGAGCATCCGACTAT 5926  
 141 S V A D T L R Q K V E D A V R E H P D Y -  
 5927 CGCGTGGTGTACCGGACATAGCTGGTGGCATGGCAACTGTTGCCGGAGGAGAC 5986  
 161 R V V F T G H S L G G A L A T V A G A D -  
 5987 CTGCGTGGAAATGGGTATGATATCGACGTGTTTCAATATGGCGCCCCCGAGTCGGAAAC 6046  
 181 L R G N G Y D I D V F S Y G A P R V G N -  
 6047 AGGGCTTTGCAGAATTCTGACCGTACAGACCGGGGAAACACTCTACCGCATTACCCAC 6106  
 201 R A F A E F L T V Q T G G T L Y R I T H -  
 6107 ACCAATGATATTGTCCTAGACTCCCGCCGGGAATTGGTTACAGCCATTCTAGCCCA 6166  
 221 T N D I V P R L P P R E F G Y S H S S P -  
 6167 GAGTACTGGATCAAATCTGGAACCTTGTCCCCGTACCCGAAACGATATCGTGAAGATA 6226  
 241 E Y W I K S G T L V P V T R N D I V K I -  
 6227 GAAGGCATCGATGCCACCGGGCGCAATAACCGCTAACATTCCGGATATCCCTGCGCAC 6286  
 261 E G I D A T G G N N Q P N I P D I P A H -  
 6287 CTATGGTACTTCGGGTTAATTGGGACATGTCTTTAG 6322  
 281 L W Y F G L I G T C L -

Fig. 1



2/10

ATGAGGGAGCTCCCTTGTGCTGTTCTTGCTCTGCGTGGACGGCCCTGGCCAGTCCTATA 5506  
 5447 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 1 M R S S L V L F F V S A W T A L A S P I -  
 SnaBI PvuII  
 CGTAGAGAGGTCTCGCAGGATCTGTTAACAGTCATCTCTTGACAGTATTCACT 5566  
 5507 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 21 R R E V S Q D L F N Q F N L F A Q Y S A -  
 GCCGCATACTGCGGAAAAAACATGATGCCAGCAGGTACAAACATTACGTGACGGGA 5626  
 5567 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 41 A A Y C G K N N D A P A G T N I T C T G -  
 SphI  
 AATGCATCCCCGAGGTAGAGAAGGCGGATGCAACGTTCTACTCGTTGAAGACTCT 5686  
 5627 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 61 N A C P E V E K A D A T F L Y S F E D S -  
 HindIII  
 GGAGTGGCGATGTCACCGGCTCCTGCTCTGACAAACACGAAACAAGCTTATCGTCCTC 5746  
 5687 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 81 G V G D V T G F L A L D N T N K L I V L -  
 BglII  
 TCTTCCGTGGCTCAAGATCTATAGAGAACTGGATCGGGATCTTAACTTCGACTTGAAA 5806  
 5747 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 101 S F R G S R S I E N W I G N L N F D L K -  
 GAAATAAAATGACATTTGCTCCGGCTGCAGGGGACATGACGGCTTCACTCGTCTGGAGG 5866  
 5807 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 121 E I N D I C S G C R G H D G F T S S W R -  
 NruI  
 TCTGTAGCCGATACTTAAAGCAGAAGGTGAGGATGCTGTTCGCGGACATCCCGACTAT 5926  
 5867 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 141 S V A D T L R Q K V E D A V R E H P D Y -  
 BstXI NheI  
 CGCGTGGTGTACCGGCCATAGCCTGGTGGCTAGCAACTGTTGGGAGCAGAC 5986  
 5927 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 161 R V V F T G H S L G G A L A T V A G A D -  
 BstEII  
 CTGCGTGGAAATGGGTATGATATCGACGTGTTTCATATGGCGCCCCCGAGTCGGTAC 6046  
 5987 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 161 L R G N G Y D I D V F S Y G A P R V G N -  
 KpnI  
 CGTGCCTTTGGAGAATTGACCGTACAGACCGGGGTACCGTACCGGATTACCCAC 6106  
 6047 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 201 R A F A E F L T V Q T G G T L Y R I T H -  
 XbaI  
 ACCAATGATATTGCTCCCTAGACTCCGCTCGAGATTGGTTACAGCCATTCTAGCCCA 6166  
 6107 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 221 T N D I V P R L P P R E F G Y S H S S P -  
 SpeI  
 GAGTACTGGATCAATTCTGGAACACTAGTCCCCGTACCCGAAACGATATCGTGAAGATA 6226  
 6167 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 241 E Y W I K S G T L V P V T R N D I V K I -  
 GAAGGCATCGATGCCACCGGGCGGAAATAACAGGCTAACATTCCGGATATCCCTGGCAC 6286  
 6227 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 261 E G I D A T G G N N Q P N I P D I P A H -  
 CTATGGTACTTCGGGTTAATTGGGACATGTTAG 6322  
 6287 -----+-----+-----+-----+-----+-----+-----+-----+-----+  
 281 L W Y F G L I G T C L -

Fig. 2



3/10

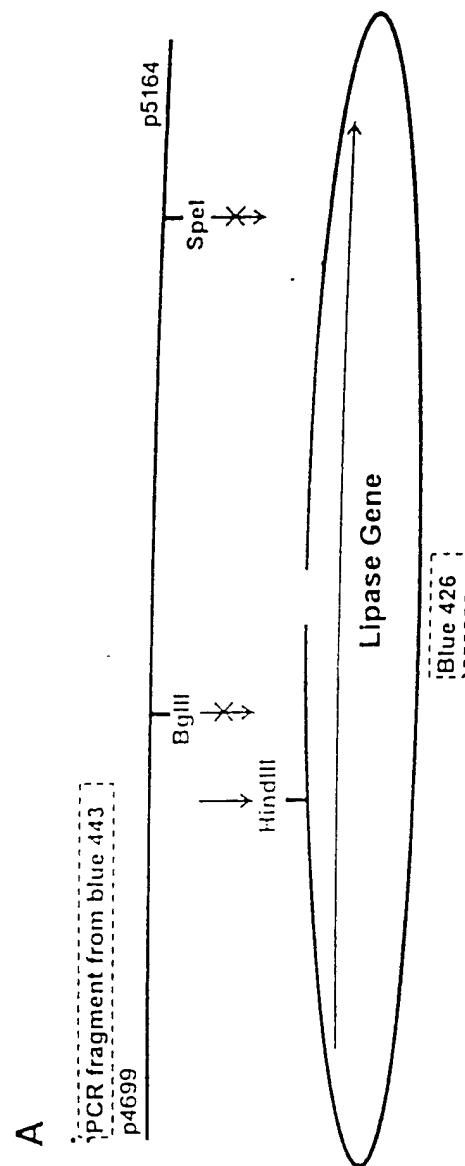
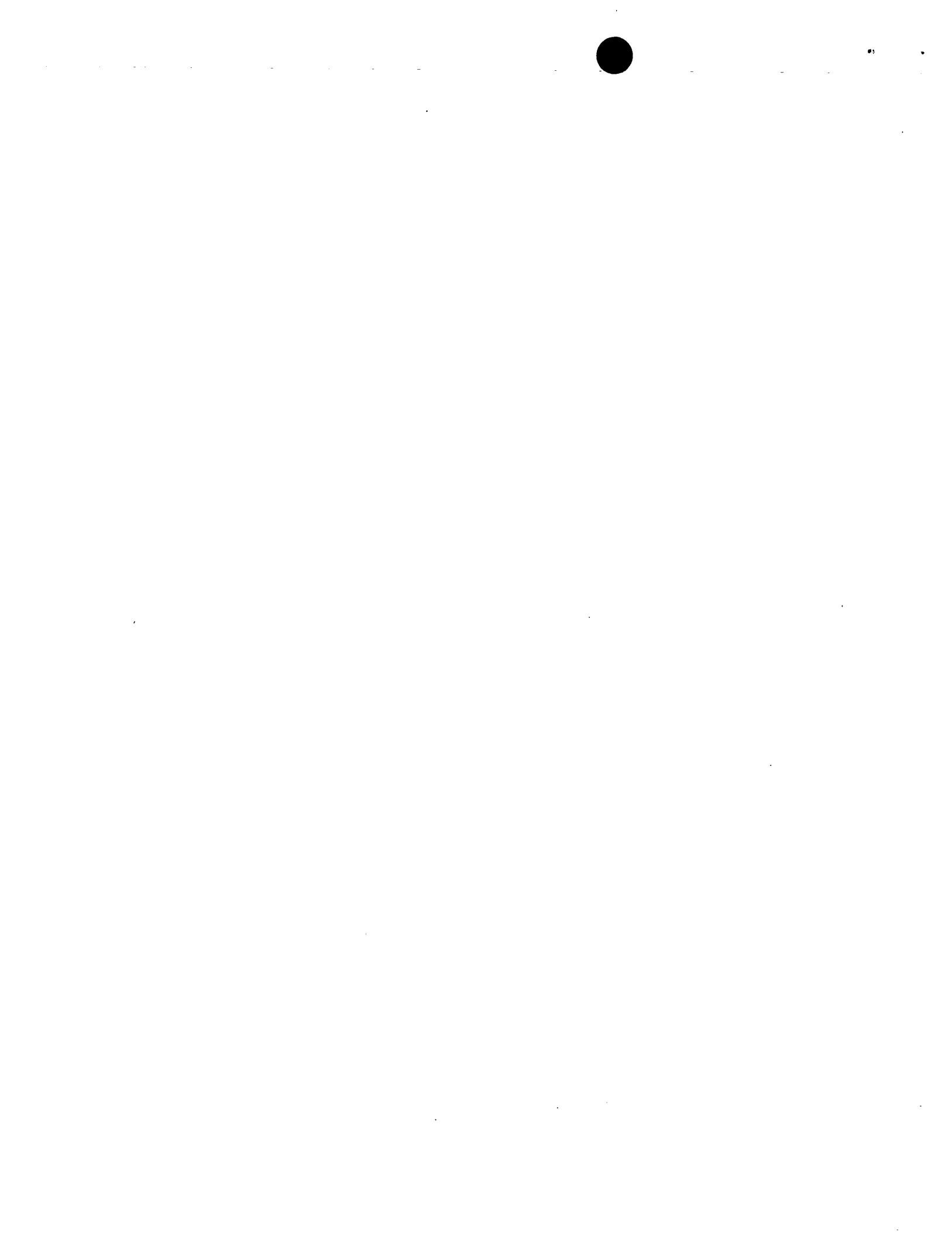


Fig. 3



4/10

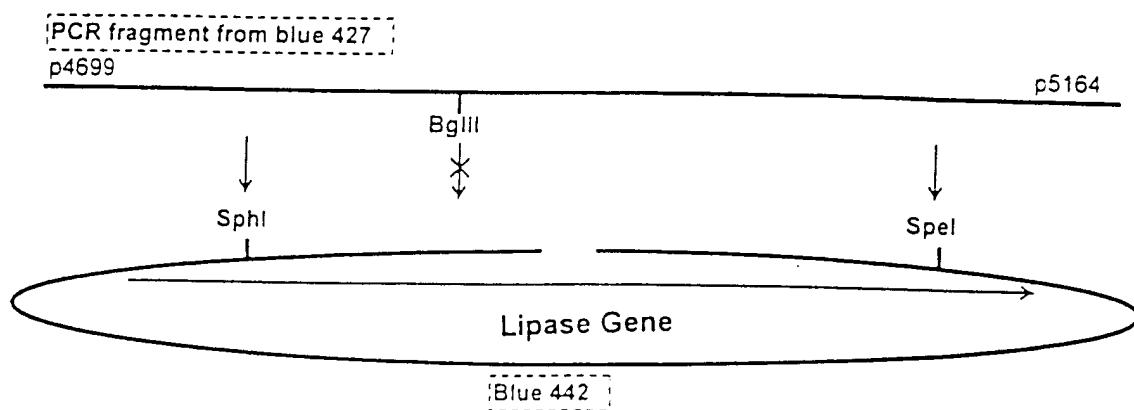
**B:**

Fig. 4



5/10

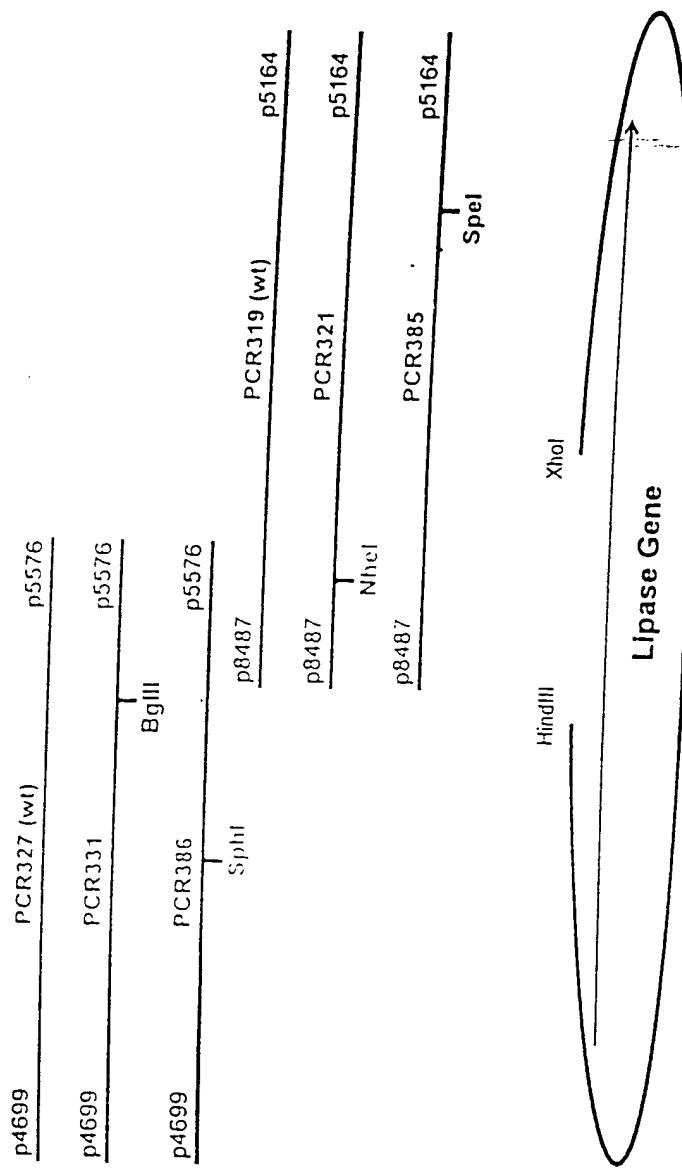


Fig. 5



6/10

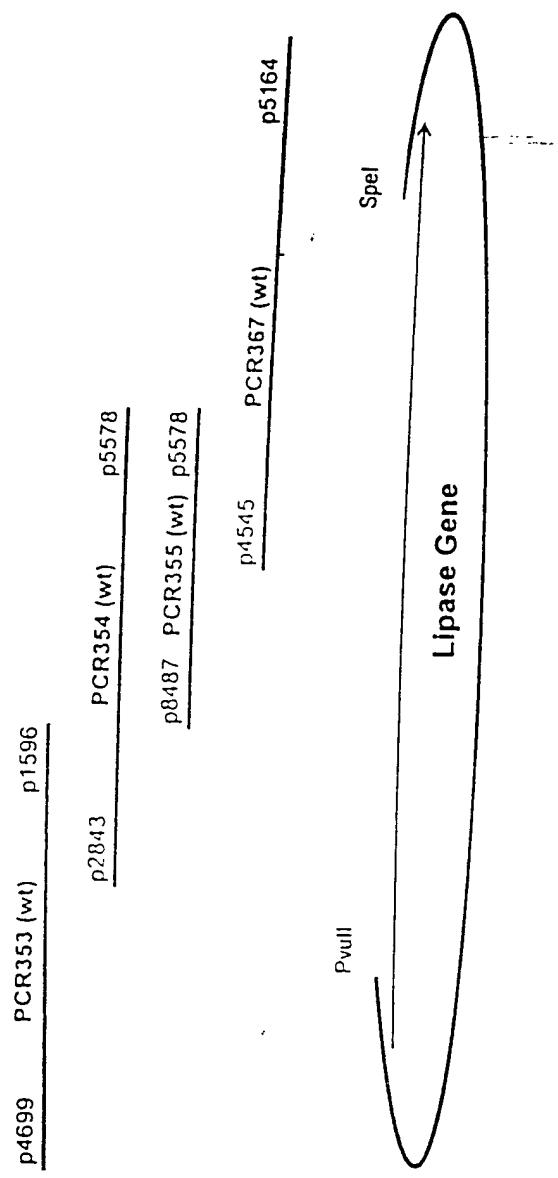


Fig. 6



7/10

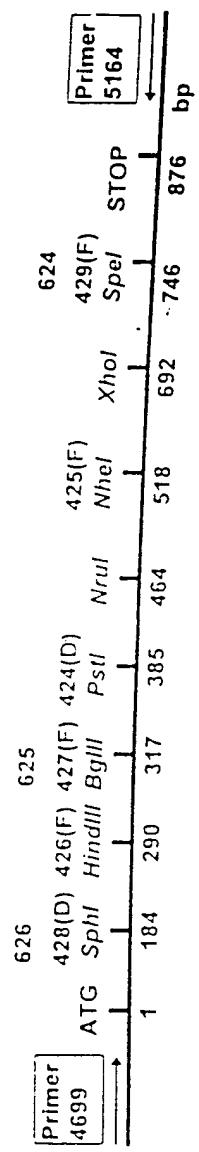


Fig. 7



8/10

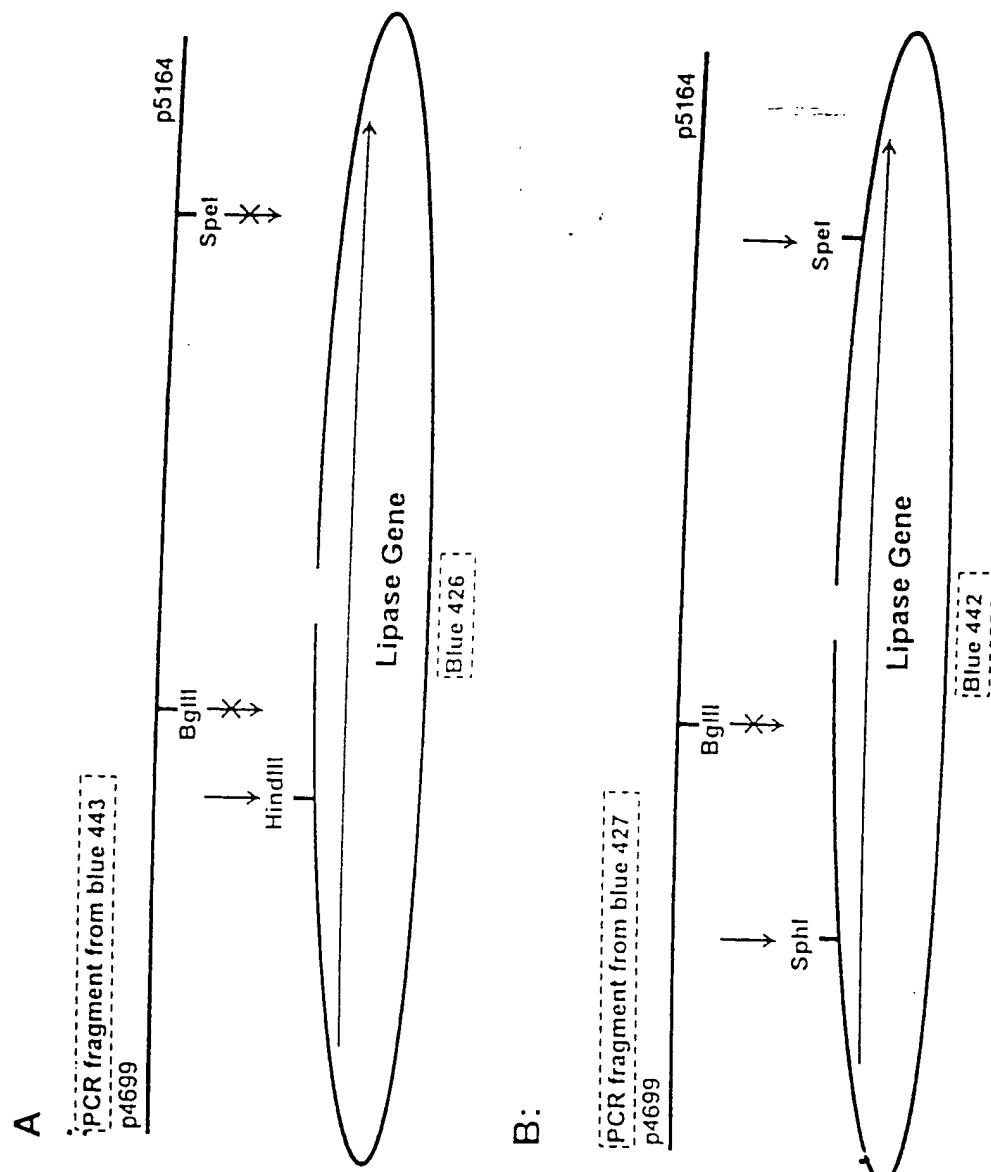


Fig. 8



9/10

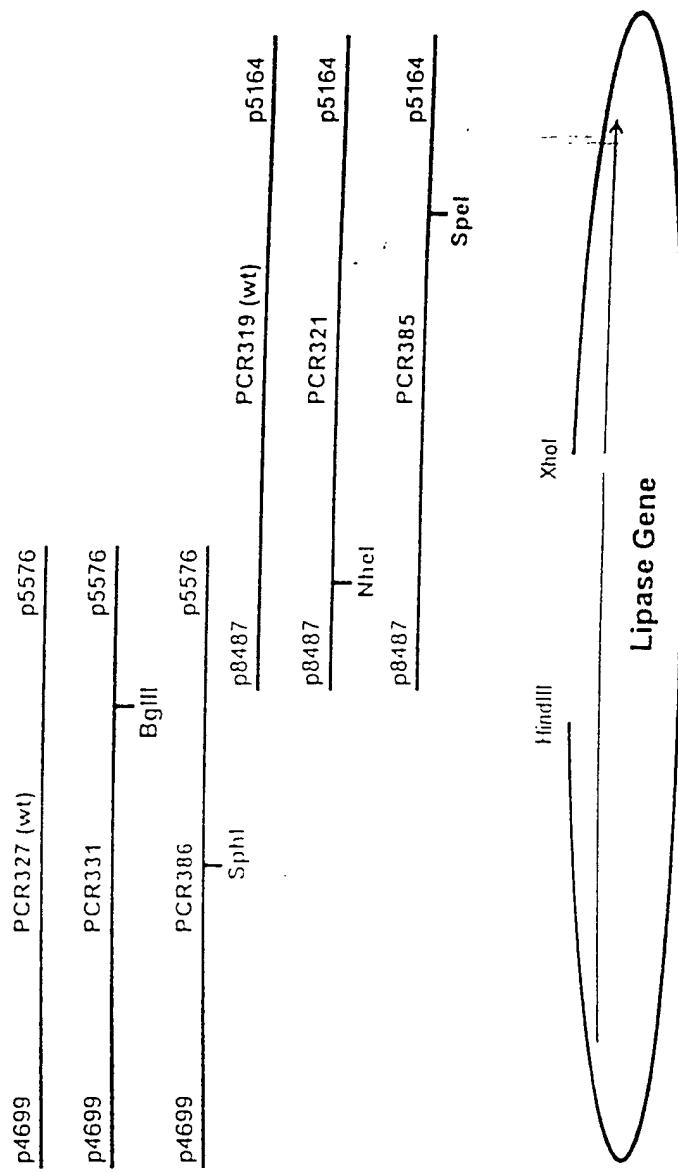


Fig. 9



10/10

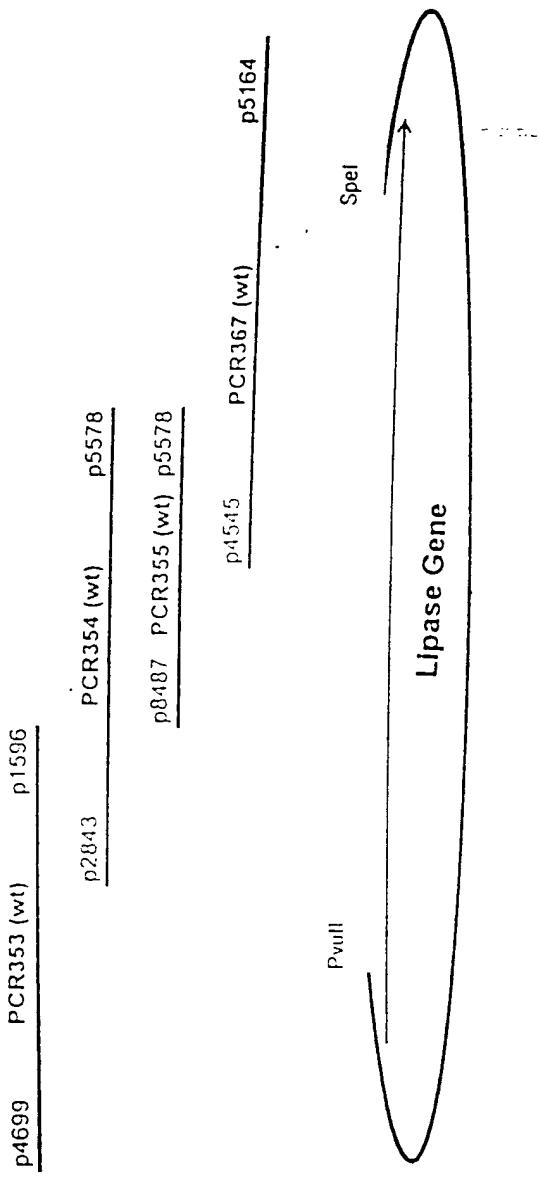


Fig. 10



## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/DK 96/00343

A. CLASSIFICATION OF SUBJECT MATTER		
<p>IPC6: C12N 15/10, C12N 15/62, C12N 15/63 According to International Patent Classification (IPC) or to both national classification and IPC</p>		
B. FIELDS SEARCHED		
<p>Minimum documentation searched (classification system followed by classification symbols)</p> <p>IPC6: C12N</p>		
<p>Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched</p> <p>SE, DK, FI, NO classes as above</p>		
<p>Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)</p> <p>WPI, EPODOC, US PAT FULL, MEDLINE, BIOSIS, DBA CA, SCISEARCH, PATENT CITATION INDEX</p>		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Gene, Volume 83, 1989, Denis Pompon et al, "Protein engineering by cDNA recombination in yeasts: shuffling of mammalian cytochrome P-450 functions", page 15 - page 24, figures 1,2 --	1-26
A	EP 0141484 A2 (BIOGEN N.V.), 15 May 1985 (15.05.85), figures --	1-26
A	US 5093257 A (GREGORY L. GRAY), 3 March 1992 (03.03.92), column 3, line 11 - line 13 -- -----	1-26
<input type="checkbox"/> Further documents are listed in the continuation of Box C.		<input checked="" type="checkbox"/> See patent family annex.
<p>* Special categories of cited documents</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&amp;" document member of the same patent family</p>		
Date of the actual completion of the international search		Date of mailing of the international search report
8 November 1996		13.11.96
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86		Authorized officer  Patrick Andersson Telephone No. +46 8 782 25 00

## INTERNATIONAL SEARCH REPORT

Information on patent family members

01/10/96

International application No.

PCT/DK 96/00343

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP-A2- 0141484	15/05/85	JP-A-	60070083	20/04/85
US-A- 5093257	03/03/92	AU-A-	5945086	08/01/87
		CA-A-	1312836	19/01/93
		DE-D, T-	3688920	23/12/93
		EP-A, B-	0208491	14/01/87
		SE-T3-	0208491	
		IE-B-	59875	20/04/94
		JP-B-	8029091	27/03/96
		JP-A-	62083891	17/04/87